

Intelligentní dopravní systémy

Intelligent Traffic Systems

Prohlašuji, že jsem tuto diplomovou práci vypracoval samostatně. Uvedl jsem všechny literární prameny a publikace, ze kterých jsem čerpal.

V Šumperku 6. května 2011

.....

Rád bych na tomto místě poděkoval vedoucímu práce Ing. Michalovi Radeckému za cenné připomínky, pomoc a rady během vedení této práce.

Abstrakt

Tato práce se zabývá analýzou dopravních dat, jakož i predikcí dopravy na základě těchto dat. Popisuje jednotlivé možnosti předpovídání rychlostí vozidel včetně vzájemného srovnání prediktivních metod, které je provedeno na reálných naměřených datech. Vybrané metody analýzy a predikce jsou implementovány s ohledem na napojení na další části systému pro podporu analýzy a simulací v oblasti dopravy.

Klíčová slova: časová řada, doprava, lineární regrese, neuronové sítě, predikce

Abstract

This thesis deal with the analysis of traffic data and traffic forecasts based on this data. It describes the various options for forecasting the speed of vehicles, including a comparison of predictive methods which is done on real measured data. Selected methods of analysis and forecasting are implemented with regard to links to other parts of the system for analysis and simulation in the field of traffic.

Keywords: forecast, linear regression, neural networks, time series, traffic

Seznam použitých zkratek a symbolů

ANN	– Artificial neural network, umělá neuronová síť
BTS	– Base Transceiver Station, Základnová převodní stanice
Floreon	– Floods Recognition On the Net
GPS	– Global Positioning System, Globální polohový systém
GSM	– Global System for Mobile Communications (dříve Groupe Spécial Mobile), Globální systém mobilní komunikace
IQR	– Interquartile range, mezikvartilové rozpětí
JSDI	– Jednotný systém dopravních informací pro ČR
RDS-TMC	– Radio Data System - Traffic Message Channel, System rádiových dat - Kanál dopravních zpráv
UMTS	– Universal Mobile Telecommunications System, Univerzální mobilní telekomunikační systém

Obsah

1	Úvod	2
2	Dopravní data	3
2.1	JSDI	3
2.2	Dopravní portál města Ostrava	4
2.3	Floreon+	5
2.4	Sběr dopravních dat	7
3	Data pro analýzu	9
3.1	Analýza dopravních dat	10
4	Vybrané nástroje predikce	12
4.1	Časová řada	12
4.2	Metody predikce	14
5	Predikce rychlostí vozidel	22
5.1	Aritmetický průměr dat	22
5.2	Lineární regrese	22
5.3	Neuronové sítě	23
6	Experimenty s prediktivními metodami	25
6.1	Aritmetický průměr dat	25
6.2	Lineární regrese	26
6.3	Neuronové sítě	27
6.4	Reálné příklady predikce	29
6.5	Zhodnocení metod predikce	33
7	Implementace a propojení s webovou aplikací	35
7.1	Implementace	35
7.2	Propojení	37
7.3	Webová aplikace	38
8	Vize do budoucna	41
9	Závěr	43
10	Reference	44

1 Úvod

Tato práce se věnuje problematice analýzy a zpracování reálných dopravních dat. Jelikož každého z nás doprava ovlivňuje, ať přímo či nepřímo, je vhodné vědět, jak se dopravní situace na různých místech právě vyvíjí a jak by se mohla chovat v budoucnu. Na základě těchto znalostí mohou řidiči dostávat do svých autonavigací s dostatečným předstihem informace například o tom, že cesta, kterou chtějí použít, bude s největší pravděpodobností velmi pomalu průjezdná a zároveň jim může být doporučena alternativní silnice s bezproblémovou průjezdností. Další možnost využití se nabízí při plánování dopravy u velkých sportovních nebo kulturních akcí či v oblasti krizového řízení.

K vytvoření takovýchto „znalostí“ musíme nejprve získat nějaká reálná dopravní data, která podrobíme důkladné analýze. Zjistíme, jakými způsoby lze taková data získávat a následně využijeme existující systém Floreon+ Traffic a údaje z něj se popíší a dále zpracují. Také se představí současné systémy, které dopravní data využívají a pracují s nimi.

Podstatná část práce je věnována předpovídání rychlostí vozidel. Nejprve se seznámíme s některými obecnými způsoby predikce (lineární regrese, neuronové sítě, použití aritmetického průměru). Ty si teoreticky představíme a ukážeme, jak s nimi můžeme pracovat. Poté se upraví a naimplementují pro použití s dopravními daty. Hotové prediktivní metody se na větším množství dopravních dat otestují a provede se jejich zhodnocení vzhledem k přesnosti předpovídání rychlosti vozidel.

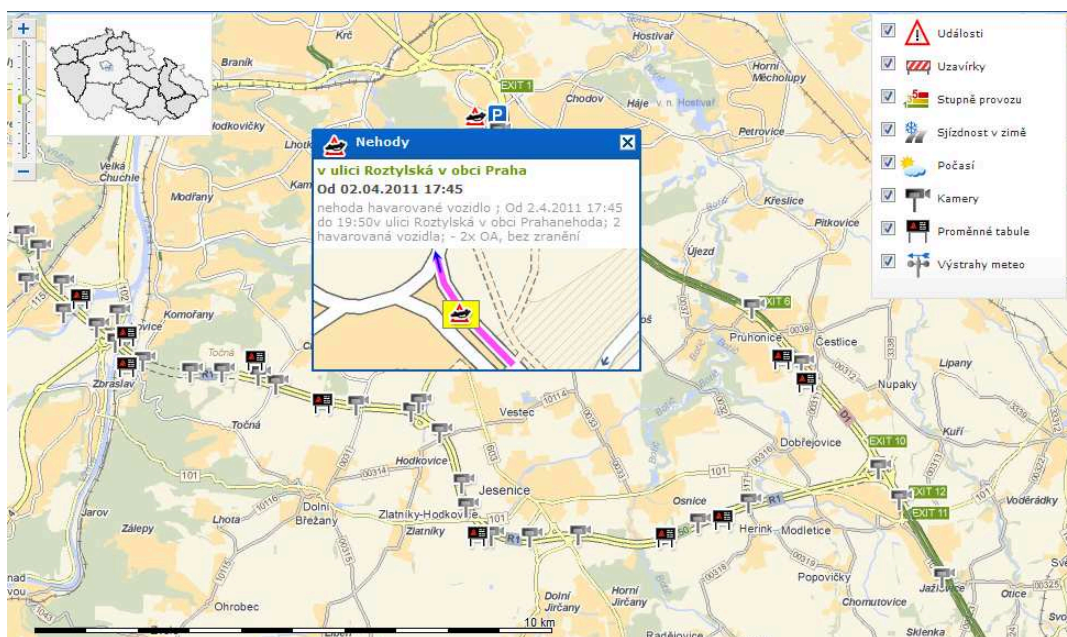
Navržené a naimplementované techniky predikce a analýzy dat se zpřístupní pro použití v rámci rozsáhlejšího systému, jehož součástí je i tato diplomová práce. Zmíníme se i o webové aplikaci, která výstupy z práce využívá ke své činnosti. Na závěr zkusíme navrhnout možná vylepšení analýzy a predikce do budoucna.

2 Dopravní data

V dnešní uspěchané době a při přeplněnosti silnic je naprosto nezbytné, aby byla doprava nějakým způsobem řízena. Nestačí jenom sledovat kamerový provoz na vybraných křižovatkách a podle toho například upravovat pouze časování semaforů. Řidiči vozidel musí být s dostatečným předstihem informováni o problematických úsecích silnic, které bývají v určitou dobu pomalu průjezdné (na základě různých analýz) a stejně tak by měli mít informace o tom, jakou trasu mají zvolit, aby se nevhodným úsekům vyhnuli. Nezbytným doplňkem takovýchto dat jsou i údaje o aktuální situaci, jako jsou dopravní nehody, uzavírky silnic, počasí a v zimě i údaje o sjízdnosti.

2.1 JSDI

Aktuální dopravní data již zpracovává Jednotný systém dopravních informací pro Českou republiku, který je také zobrazuje na mapě dostupné na internetových stránkách <http://www.dopravniinfo.cz> (obrázek 1) a prostřednictvím vysílačů Českého rozhlasu (Radiožurnál a Vltava) jsou k dispozici i v rámci služby RDS-TMC [10] pro příjem v navigacích, které je mohou zahrnout do svých plánovačů tras. JSDI je společným projektem Ministerstva vnitra ČR, Ministerstva dopravy ČR, Ředitelství silnic a dálnic ČR a dalších veřejných a soukromých institucí.



Obrázek 1: Mapa JSDI s aktuálními dopravními informacemi

JSDI shromažďuje data z agendových systémů především veřejné správy (dopravní nehody, požáry, havárie, údržba silnic, parkování, uzavírky, nadměrné náklady, velké akce), z telematických systémů (dohledové kamery, počasí, elektronické myto, řízení

tunelu, liniové řízení, sčítání dopravy, detekce kolon) a také přímo od řidičů. Tato data po obdržení zpracovává Národní dopravní informační a řídicí centrum a jsou dále šířena dalším systémům [11].

Po zpracování se objeví ve výstupu jako nehody, uzavírky, překážky v provozu, práce na silnici, stupně provozu na komunikaci, zimní sjízdnost nebo počasí. Dále je možné zobrazit i aktuální obrázky z kamer nebo texty z proměnných tabulí. Bohužel některé z informací jsou dostupné pouze na dálnicích a vybraných rychlostních silnicích (stupně provozu, proměnné tabule, kamery).

2.2 Dopravní portál města Ostrava

Statutární město Ostrava nabízí na svých webových stránkách <http://doprava.ostrava.cz> interaktivní dopravní portál (obrázek 2), který umožňuje na mapě Ostravy (využívá mapy poskytované společností Google) zobrazovat tři druhy dopravních informací – o dopravě, o počasí a o parkování.



Obrázek 2: Dopravní portál města Ostrava s vyznačenými parkovišti

Informace o dopravě jsou přebírány z JSDI a zahrnují dopravní omezení a nehody. Zajímavou funkcí je dopravní situace, která upozorňuje na chystané akce ve městě, které by mohly ovlivnit provoz v některé ulici. Údaje o počasí obsahují data o oblačnosti, srážkách, větrnosti, viditelnosti a stavu povrchu vozovky. Bohužel se na mapě vykreslují pouze v zimním období (doba od 1. listopadu do 31. března), i když by je řidiči v Ostravě určitě uvítali celoročně.

Dopravní portál také nabízí informace o parkování, které jsou zatím dostupné pro centrum Ostravy. Na mapě zobrazuje pojmenované parkovací oblasti, jejichž součástí jsou i vybraná parkoviště. U nich jsou údaje o kapacitě parkoviště, aktuálním počtu

volných míst a fotografie areálu. K dispozici je i odkaz vedoucí na bližší informace o konkrétním parkovišti. K parkovištím se vztahují i navigační tabule umístěné v centru města a sdělující řidičům, kolik volných míst nabízí parkoviště a jakým směrem se k nim dostanou. Poslední volbou jsou parkovací zóny a parkovací hodiny ukazující název zóny a cenu za parkování.

2.3 Floreon+

Projekt Floreon+ je výsledkem budování systému pro modelování a simulace situací v Moravskoslezském kraji, které jsou způsobeny nepříznivými přírodními jevy, přičemž se využívají moderní internetové technologie. Ze začátku se v projektu řešilo modelování, simulace a predikce povodní. Jelikož se tento povodňový systém osvědčil, začali se budovat jeho další části. V současnosti se skládá z těchto modulů a základních funkcí [12]:

- Voda
 - Podpora povodňového krizového řízení
 - Rozšíření možností předpovídání
 - Grafické znázornění povodňových prognóz
 - Problematika přívalových srážek a předpovídání vývoje bleskových povodní (v řešení)
- Doprava
 - Vývoj dopravní situace v průběhu krizových událostí
 - Předpovídání dopravních omezení v průběhu krizových událostí (v řešení)
- Kvalita životního prostředí (v řešení)
 - Krátkodobá předpověď vývoje kvality ovzduší
 - Předpovídání vývoje smogu a inverzí
 - Simulace úniku znečišťujících látek do ovzduší a vody

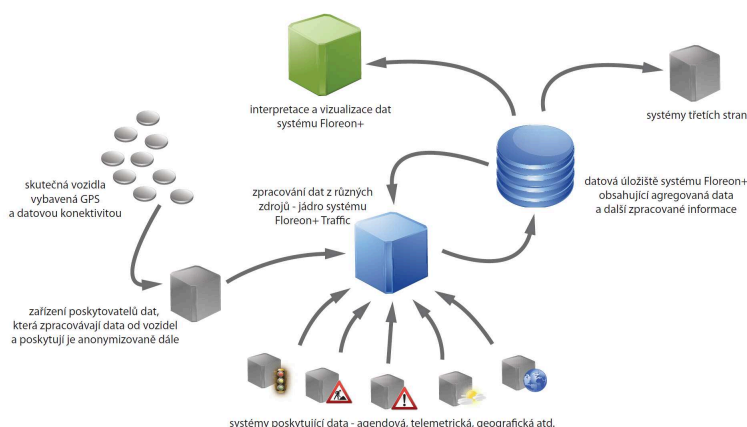
Systém Floreon+ je primárně určen složkám integrovaného záchranného systému, ale s několika určitými omezeními ho může používat i široká veřejnost a běží na adrese <http://floreon.vsb.cz>.

Celý systém se skládá z několika vrstev [13]. Nejspodněji jsou umístěna datová uložiska skládající se především z databázových serverů PostgreSQL a Microsoft SQL Server. Ve druhé vrstvě leží různé výpočetní moduly, moduly pro dopravu a znečištění. Další vrstva se stará o řízení celého systému. Poslední částí jsou koncové programy a aplikace pro komunikaci s jinými systémy. Pro tuto práci je klíčová „dopravní část“, která bude později využívána a proto si ji blíže popíšeme.

2.3.1 Floreon+ Traffic

Jak již bylo zmíněno Floreon+ Traffic je nedílnou součástí celého systému a vhodně tak doplňuje povodňový základ projektu. Tento modul získává data z různých zdrojů vztahujících se k dopravě a vytváří z nich agregovaná data určená pro další využití (obrázek 3). Mezi takovéto zdroje patří:

- reálné údaje o rychlosti a lokaci z vozidel zapojených do projektu
- data z JSDI a telematických systémů
- data z kamerových systémů
- informace o počasí

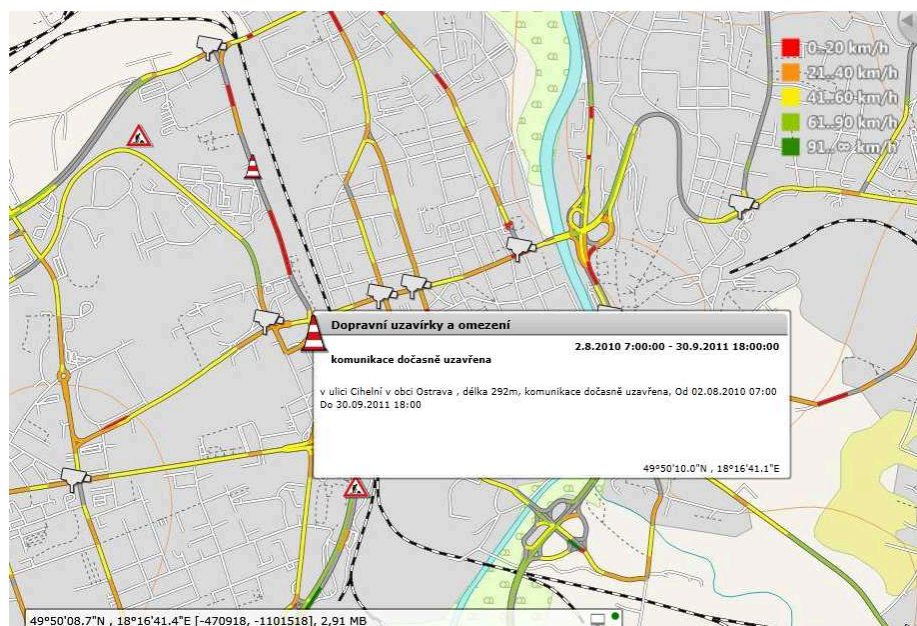


Obrázek 3: Schéma modulu Floreon+ Traffic [14]

přičemž zásadní pro systém jsou první dva zdroje. JSDI je popsán v předchozí podkapitole, takže přejdeme rovnou k reálným údajům o rychlosti. Tyto data jsou získávána z vozidel zapojených do projektu a pohybujících se ve sledované oblasti a obsahují údaje o poloze vozidla, jeho rychlosti a času pořízení záznamu. Obdržené informace jsou anonymizované, takže z nich není možné zpětně zjistit vozidlo, ke kterému se záznam vztahuje. Souřadnice polohy získané z GPS jsou speciálním algoritmem převedeny na konkrétní úsek cesty. V roce 2009 bylo zapojeno 2200 vozidel a ve špičce generovaly 13 záznamů za sekundu, přičemž ve finále je použito asi 30 % získaných záznamů a ty jsou po hodinách agregovány pro každý úsek cesty [14].

Jakmile je nasbírán dostatek dat, lze začít vytvářet informace o dopravní situaci. Kromě existujících agregovaných dat, která se mohou dále zpracovávat, poskytuje systém zejména:

- statistické informace o úsecích
- vizualizaci dat (obrázek 4), jako jsou třeba nehody, kamerové systémy, stupně provozu, uzavírky



Obrázek 4: Mapa Floreonu+ s aktuální dopravní situací

2.4 Sběr dopravních dat

Sběr dopravních dat je alfou a omegou každého dopravně-informačního systému. Floreon+ Traffic využívá především data, která poskytují spolupracující vozidla. Samozřejmě to není jediný možný způsob sběru dopravních dat a do budoucna ani příliš efektivní, pokud by nebyl doplněn o další možnosti. Techniky získávání dopravních dat lze rozdělit do následujících kategorií:

- profilová měření
 - detektory
 - mýtný systém
 - úsekové měření
 - kamery
- plovoucí vozidla využívající GPS/GSM
- informace z BTS

Z profilových měření se u nás nejvíce využívá profilových detektorů na bázi elektromagnetické smyčky, které jsou vloženy do povrchu vozovky [15]. Dalšími variantami může být úsekové měření rychlosti nebo rychlostní kamery provozované Policií ČR případně městskou (obecní) policií. Dopravní data již sbírá i mýtný systém pomocí mýtných bran, ta ale nejsou nějak speciálně využívána. Mýtné brány momentálně evidují pouze nákladní

auta, ale s plánovaným rozšířením i na osobní vozy by byla k dispozici zajímavá dopravní data, pokrývající všechny hlavní dopravní tahy v České republice.

Plovoucí vozidla využívající GPS jsou vybraná vozidla, která v určitých intervalech posílají prostřednictvím mobilní datové sítě (GSM/UMTS) informace o vozidle a jeho rychlosti (tento způsob je využíván v systému Floreon+ Traffic). Tyto informace jsou anonymizovány a dále zpracovávány. Plovoucí vozidla mohou být založena i na signalizačních datech mobilních sítí (GSM/UMTS) [15] neboli komunikaci mobilních telefonů se základnovými stanicemi (BTS) buňkové sítě. Tato komunikace ovšem může být využita i mimo běžná vozidla a umožňuje sledovat přesun velkého počtu osob (např. v městské hromadné dopravě, při konání sportovních či kulturních akcích).

Zatímco využití profilových detektorů a plovoucích vozidel s GPS je v podstatě bezproblémové, u ostatních technik už to neplatí, protože nejsou anonymní a jejich sběr probíhá bez souhlasu osob, které řídí nebo vlastní zaznamenané vozidlo. Takovýto sběr dat by pak mohl být v rozporu se zákony, obzvlášť s normami věnující se ochraně osobních údajů¹.

¹Zákon č. 101/2000 Sb. o ochraně osobních údajů

3 Data pro analýzu

V předchozí části byl představen modul Floreon+ Traffic, který mimo jiné shromažďuje agregovaná dopravní data. Protože z těchto záznamů budou vycházet analytické a prediktivní metody, představíme si je blíže. Každý dopravní záznam obsahuje následující atributy:

- číslo úseku cesty
- čas
- průměrná rychlost
- počet vozidel

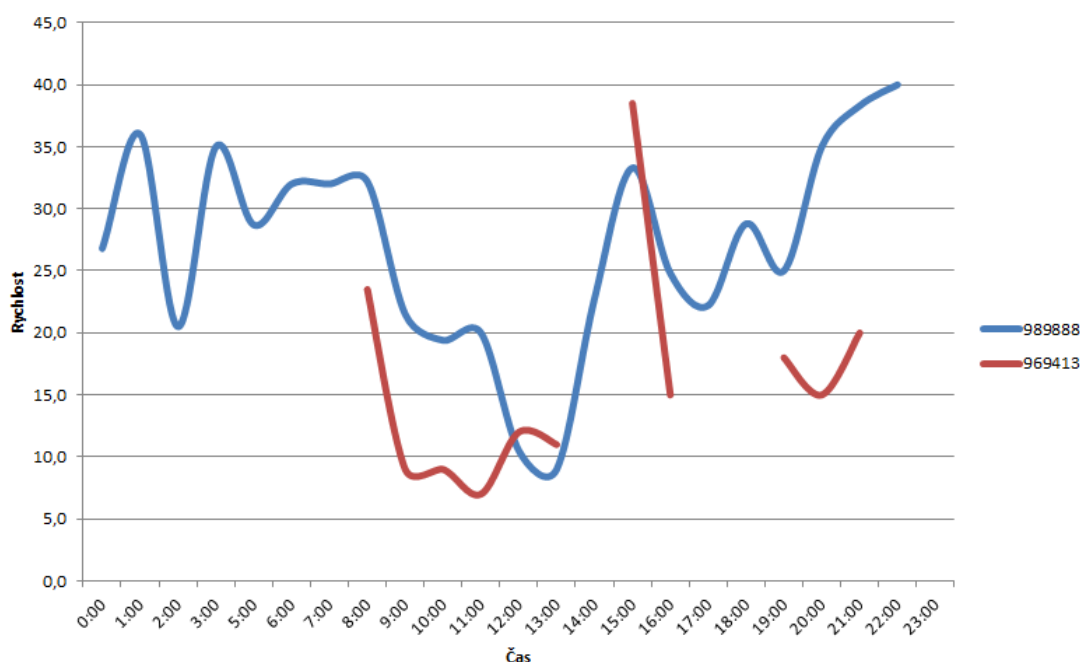
Číslo úseku cesty (ID) představuje interní identifikátor, který má vazbu na identifikátor cest používaný v mapě. Čas vyjadřuje dobu, ke které byla data agregována; agregace se provádí vždy k celé hodině. Průměrná rychlost je průměrnou rychlostí všech vozidel, které projely daným úsekem cesty během hodiny. Počet vozidel značí celkový počet vozidel projelých úsekem komunikace během hodiny. Tabulka 1 ilustruje několik takových záznamů. Z této tabulky je možné vidět výhodu i nevýhodu těchto dat. Poslední dva

Úsek cesty	Čas	Rychlost	Počet vozidel
718596	30. 10. 2010 12:00	85	1
718596	15. 10. 2010 8:00	80	4
718596	25. 11. 2010 6:00	15,2	5
558467	14. 1. 2011 11:00	25,5	297
730987	14. 1. 2011 10:00	7,7	362

Tabulka 1: Agregované dopravní záznamy

záznamy jsou agregovány z poměrně vysokého počtu vozidel (5 - 6 vozidel za minutu), které úsekem projely během dané hodiny, a lze s nimi celkem spolehlivě pracovat. První tři úseky se naopak vyznačují velmi nízkým počtem projelých vozidel a vypočítaná průměrná rychlost tak vůbec nemusí odpovídat skutečnosti. S tím souvisí i další negativum, které spočívá v tom, že u spousty sledovaných úseků mnohdy neprojde ani jedno auto za den.

Dalším problémem může být to, že vozidla zapojená do měření jsou z kategorie pracovních vozů a tomu odpovídají i časy, ve kterých jezdí po komunikacích. Ty obvykle kopírují pracovní dobu jednotlivých firem a tak je k dispozici poměrně málo údajů z večerních nebo nočních hodin. Nejzávažněji se tento stav projevuje o víkendu, kdy je počet projelých vozidel tak nízký, že nemá příliš smysl tato data dále analyzovat. Například o víkendu 26. a 27. března 2011 bylo evidováno 1348 záznamů, z toho na nejvytíženějších úsecích projelo 6 vozidel za hodinu (v sedmi případech) a průměrný počet vozidel za hodinu se rovnal jednomu. Oproti tomu v pracovní den v pondělí 14. března 2011 databáze eviduje 40 867 záznamů.



Obrázek 5: Graf vývoje rychlostí během dne

Pro lepší ilustraci těchto „vhodných“ a „nevhodných“ dat je na obrázku 5 znázorněn graf s vývojem rychlostí vozidel během jednoho dne (14. března 2011) na dvou úsecích (989888, 969413). Zatímco první úsek má celý den v podstatě pokrytý, druhý už je pokryt z necelé poloviny. I když druhý úsek obsahuje pouze polovinu dat, které by mohl mít, jsou to ještě dobrá čísla. V průběhu dne totiž vozidla navštívila 20 802 úseky, ale 14 036 z nich mělo za den pouze jeden záznam, 3 636 z nich dva a 1 452 tři záznamy (celý přehled dne ukazuje tabulka 2).

Počet hodinových záznamů	1	2	3	4	5	6	7	8	9
Počet úseků cest	14 036	3 636	1 452	652	365	238	123	96	52
Počet hodinových záznamů	10	11	12	13	14	15	16	17	18
Počet úseků cest	43	30	19	19	6	8	7	5	2
Počet hodinových záznamů	19	20	21	22	23	24			
Počet úseků cest	2	3	2	3	2	1			

Tabulka 2: Počet úseků vzhledem k hodinovým záznamům

3.1 Analýza dopravních dat

V rámci této práce bylo implementováno několik metod, které zpřístupňují následující funkce:

- časová řada (soubor hodnot čas/rychlost vozidel)
- základní statistika (počet vozidel, minimální a maximální rychlosti vozidel, střední hodnota, průměr a medián rychlosti vozidel, špičatost a šikmost rychlostí a jejich rozptyl a směrodatnou odchylku)
- průměrná rychlost vozidel očištěná o extrémní pozorování
- predikce rychlostí vozidel

Všechny tyto metody pracují s konkrétním úsekem cesty a navíc se dají časově omezit (od-do). Funkce základní statistiky jsou podrobně popsány v [4] a [1]. Problematice odlehlých (extrémních) pozorování a predikcí jsou věnovány následující kapitoly.

4 Vybrané nástroje predikce

Predikce nebo také předpovídání se snaží na základě určitých hodnot jevu z minulosti odhadnout možný vývoj tohoto jevu v budoucnosti. S předpovídáním se setkáváme dnes a denně při nejrůznějších činnostech našeho života. Asi nejčastěji to bude předpověď počasí, dále například demografický vývoj populace, pohyby jednotlivých položek na burze nebo určování vývoje různých přírodních katastrof. I z těchto několika málo konkrétních činností je také zřejmé, že bohužel ne vždy vede předpovídání k očekávaným výsledkům.

Abychom byli schopni vytvářet kvalitní předpovědi, není možné používat pouze obecné techniky predikce, které spočívají v různých matematických metodách a modelech, ale je nutné tyto techniky přizpůsobovat či doplňovat v závislosti na tom, co vůbec hodláme předpovídat a na jakou dobu tuto prognózu vytváříme. Vezmeme-li například předpověď počasí, je jasné, že k uspokojivé předpovědi nestačí sledovat pouze trendy jednotlivých veličin (teplota, tlak, vlhkost), ale i spoustu dalších fyzikálních hodnot, které na první pohled s počasím vůbec nemusí souviset.

Máme-li vytvořeny nějaké techniky predikce a jsme schopni pomocí nich provádět předpovědi, dají se výsledky rozdělit na dvě kategorie:

- předpověď hodnoty
- předpověď trendu

První typ výsledků předpovídá pouze nějakou hodnotu v určitý čas, např. že 14. července bude venkovní teplota 26 °C. Druhý typ predikuje pouze trend, jakým se bude sledovaný jev ubírat – klesání, stoupání, případně stejný stav.

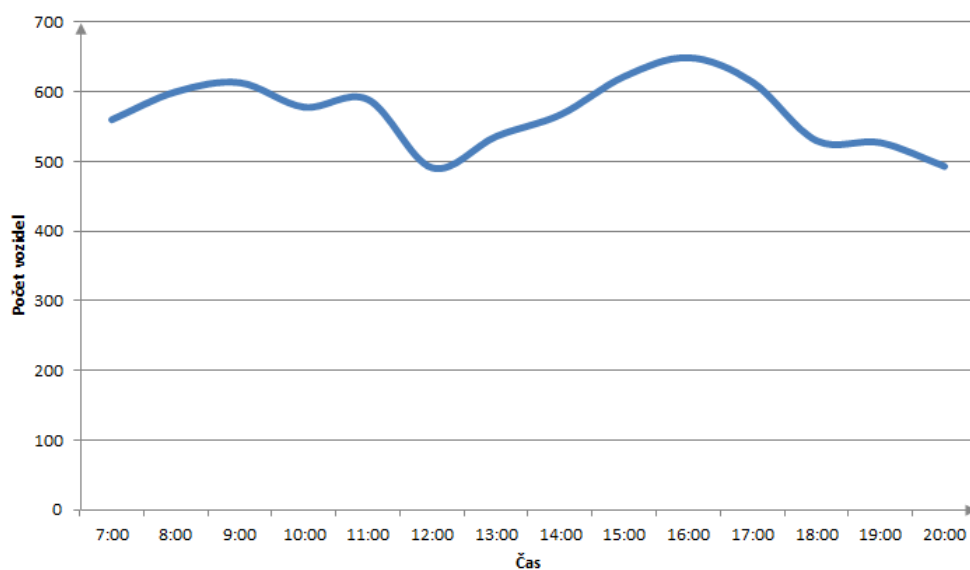
4.1 Časová řada

Časové řady vyjadřují závislost určitého sledovaného jevu na čase a to vždy od časově nejstarších hodnot po časově nejmladší hodnoty a jsou tak alfa a omega každé predikce. Máme-li k dispozici takovouto řadu, můžeme ji analyzovat (stanovovat minima, maxima, průměry,...), hledat v ní trendy nebo poslouží jako vstupní data pro další výpočty. Časová řada může graficky vypadat jako na obrázku 6 nebo ji lze vyjádřit v tabulkové formě jako tabulka 3.

Časové řady lze rozdělit na tři typy podle toho, jaký je význam časových intervalů mezi jednotlivými záznamy [2]:

1. časová řada intervalových (úsekových) ukazatelů
2. časová řada okamžikových ukazatelů
3. časová řada odvozených charakteristik

V případě prvního typu časové řady platí, že velikost ukazatele (statistického znaku) sledovaného jevu závisí na délce intervalu (měsíc, čtvrtletí, rok). U řady s okamžikovými ukazateli se hodnota plynule mění v čase a ukazatel se vztahuje k přesně definovanému



Obrázek 6: Časová řada - závislost počtu vozidel na čase

Čas	Počet vozidel
7.00	560
8.00	600
9.00	613
10.00	578
11.00	589
12.00	491
13.00	536
14.00	567
15.00	622
16.00	649
17.00	614
18.00	530
19.00	527
20.00	493

Tabulka 3: Časová řada - závislost počtu vozidel na čase

okamžiku. Při predikování se nejčastěji používá právě řada s okamžikovými ukazateli. Poslední typ časové řady je odvozován z okamžikových nebo intervalových absolutních hodnot.

4.2 Metody predikce

Pro předpovídání hodnot existuje spousta více či méně spolehlivých metod. Vždy je vhodné vybírat tyto metody podle vstupních dat (množství a kvalita dat), které máme k dispozici a také s ohledem na jev, který se bude predikovat. V této části textu se představí tři základní techniky a dojde k jejich popisu a vysvětlení. Půjde o:

- aritmetický průměr
- lineární regrese
- neuronové sítě

4.2.1 Aritmetický průměr

Ač to nemusí být na první pohled zřejmé, může aritmetický průměr sloužit jako elementární metoda pro predikci, případně se stane vstupní hodnotou pro jiné techniky předpovídání. Aritmetický průměr je jednou z nejpoužívanějších měr polohy a představuje průměrnou nebo typickou hodnotu určitého výběru [1]. Jeho hodnotu lze spočítat podle vztahu

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

kde x_i jsou jednotlivé hodnoty a n představuje počet hodnot.

Problém s použitím aritmetického průměru může nastat v okamžiku, kdy vstupní data obsahují několik extrémních hodnot, které poté průměr zkreslí a nemá tak dostatečnou vypovídací hodnotu. Toto negativum lze vyřešit identifikací a odstraněním odlehlých pozorování. K tomuto úkolu můžeme využít několik statistických metod (metody vnitřních hradeb, vnějších hradeb, z-souřadnice, mediánová souřadnice). Metody jsou podrobně popsány v [4], vzhledem k budoucímu použití se blíže zaměříme pouze na metodu vnějších hradeb.

Metoda vnějších hradeb spočívá v odstranění zvláštního případu odlehlých pozorování – extrémního pozorování. To jsou taková data, která leží mimo vnější hradby, ale nikoliv mezi vnější a vnitřní hradbou. Vnější hradby se určují za pomoci mezikvartilového rozpětí (rozdíl horního a dolního kvartilu) a samotného dolního respektive horního kvartilu. Dolní kvartil ($x_{0,25}$) rozděluje data tak, že $\frac{1}{4}$ hodnot je menší než tento kvartil a $\frac{3}{4}$ hodnot jsou větší nebo rovny tomuto kvartilu. Horní kvartil ($x_{0,75}$) rozděluje data tak, že $\frac{3}{4}$ hodnot jsou menší než tento kvartil a $\frac{1}{4}$ hodnot je větší nebo rovna tomuto kvartilu. Vyjádřeno matematicky

$$IQR = x_{0,75} - x_{0,25}$$

$$H_D = x_{0,25} - 3 \cdot IQR$$

$$H_H = x_{0,75} + 3 \cdot IQR$$

Za odlehlá (extrémní) pozorování se považují hodnoty ležící mimo interval $\langle H_D; H_H \rangle$.

Například pro hodnoty 45, 130, 60, 50, 55, 48, 61, 110, 53, 49, 5, 59, 50, 60, 10 vrátí aritmetický průměr výsledek 56,33. Aplikujeme-li na tato čísla metodu vnějších hradeb, dojde k odstranění hodnot 130, 110, 5 a 10 a následný aritmetický průměr vrátí 53,64. Rozdíl mezi těmito čísly je 2,69. K ještě většímu rozdílu (6,5) dospějeme, budeme-li mít menší množinu hodnot – 45, 135, 50, 55, 5, 52. V prvním případě dojdeme k průměru 57, v druhém po odstranění čísel 135 a 5 je výsledek 50,5.

4.2.2 Lineární regrese

Lineární regrese je další způsob predikce. Vyjadřuje aproximaci vstupních hodnot polynomem prvního řádu, v tomto případě přímkou. K tomuto úkolu využívá metodu nejmenších čtverců. Metoda nejmenších čtverců hledá takové řešení, aby minimalizovala součet druhých mocnin chyb tohoto řešení [3]. Vyjdeme tedy ze závislosti dvou proměnných – x a y . Závislost můžeme vyjádřit rovnicí $y = f(x)$. Hovoříme-li o lineární závislosti, má rovnice tvar

$$y = ax + b$$

a jedná se zároveň o vyjádření rovnice přímky, kde y označuje závislou proměnnou, x nezávislou proměnnou, koeficient a je směrnice přímky a koeficient b udává svislý posun přímky (jinak také hodnotu proměnné y , když proměnná $x = 0$).

Nyní je potřeba vyjádřit koeficienty a a b

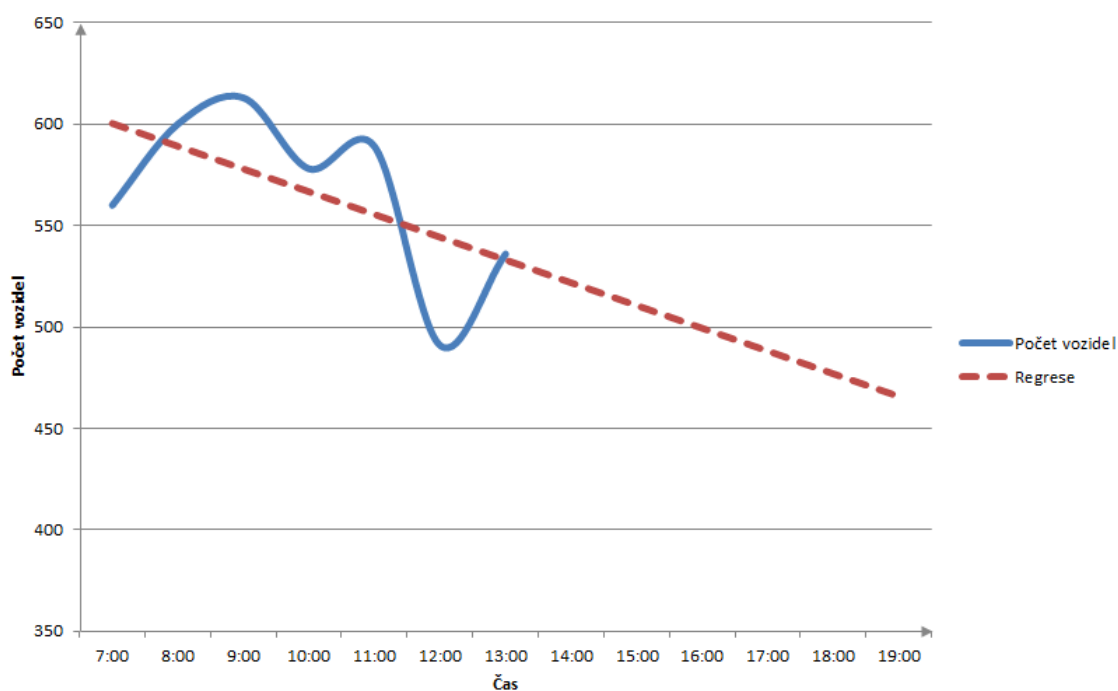
$$a = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2}$$

$$b = \frac{\sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i - \sum_{i=1}^n x_i \sum_{i=1}^n x_i y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2}$$

přičemž pro hodnotu n musí platit, že $n > 2$, jinak nelze s lineární regresí dosáhnout odpovídajících výsledků. Jakmile se na základě souboru již existujících hodnot vypočítají koeficienty a a b , je možné provádět lineární regresi. Na tomto místě je dobré podotknout, že čím více se odhadovaná hodnota vzdaluje od poslední hodnoty použité pro výpočet koeficientů a a b , tím méně je výstup z lineární regrese odpovídající.

Lineární regrese může posloužit nejenom k určení hodnoty jevu, který nastane v budoucnosti, ale také pro zobrazení regresní křivky (v našem případě přímky), která poslouží jako zobrazení možného vývoje jevu. Obrázek 7 demonstuje výstup lineární regrese, která jako vstupní data dostala údaje o počtu vozidel z horní části tabulky 4 a vytvořila z nich regresní křivku.

V dolní části tabulky 4 jsou zveřejněny pomocí lineární regrese odhadnuté hodnoty počtu vozidel na základě údajů naměřených mezi sedmou a třináctou hodinou. Z těchto čísel je dobře patrné i negativum této metody, kdy do odpoledních hodin stále předpovídá klesající počet vozidel, i když by měl mít stoupající tendenci.



Obrázek 7: Ukázka lineární regrese

Čas	Počet vozidel
7.00	560
8.00	600
9.00	613
10.00	578
11.00	589
12.00	491
13.00	536
14.00	522
15.00	511
16.00	499
17.00	488
18.00	477
19.00	466

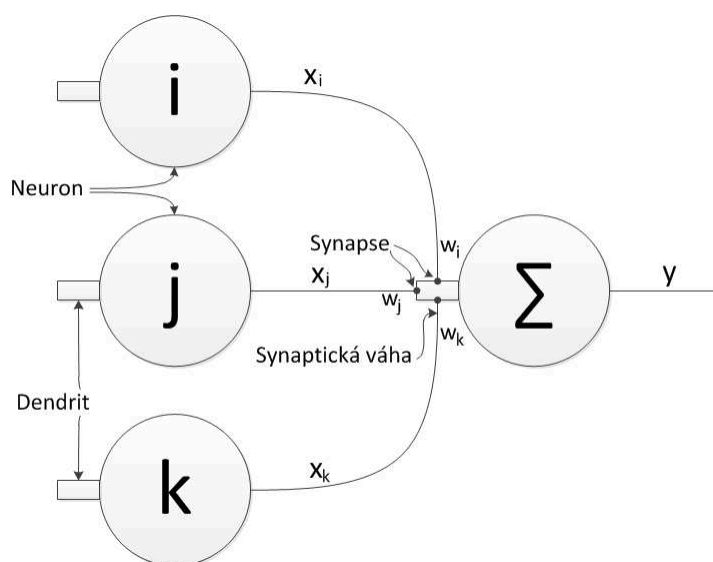
Tabulka 4: Vstupní data pro lineární regresi a predikce

4.2.3 Neuronové sítě

Poslední prediktivní metodu představují neuronové sítě. Umělé neuronové sítě jsou jednou ze základních struktur umělé inteligence a jsou inspirovány biologickými neurono-

vými sítěmi lidského mozku. Při provádění výpočtů využívají distribuované paralelní zpracování informací. Tyto sítě lze dnes využívat ve spoustě oborů - umožňují rozpoznávat obrázky, provádět kompresi dat a hlavně předpovídat vývoj časových řad.

Zásadním rozdílem mezi „běžně používanými algoritmy“ a neuronovými sítěmi je princip učení. Ten spočívá v tom, že se neuronová síť snaží na základě dodaných vstupních dat, která se označují jako trénovací množina, pochopit vazby mezi těmito daty a pak z nich správně odvozovat. Odpadá tak nutnost konstruovat složité výpočetní algoritmy pro jednotlivé problémy a nezanedbatelným přínosem je i rychlost následného dotazování, které spočívá ve sčítání a násobení relativně malého množství čísel oproti náročným výpočtům „běžných algoritmů“. Je třeba ale říci, že samotný proces učení může být v závislosti na složitosti trénovací množiny a architektury neuronové sítě časově i výkonově poměrně náročný.



Obrázek 8: „Bio/technické“ schéma neuronu

Obrázek 8 znázorňuje model neuronu, používaného v neuronových sítích v kombinaci s popisem prvků u biologických neuronových sítí. Dendrit je místo vstupu signálu do neuronu, samotný neuron slouží ke sčítání signálů přijatých od okolních neuronů, tím se stanoví vnitřní potenciál neuronu, který vede k jeho excitaci. Synapse reprezentuje výstup neuronu, kde se zeslabuje nebo zesiluje signál a pomocí axonového vlákna se předává dalším neuronům [5].

Parametry x_i, x_j, x_k značí výstupní signál svých neuronů (i, j, k), w_i, w_j, w_k jsou synaptické váhy, které upravují signály x_i, x_j, x_k a y je výstupní excitační signál. Jeden z prvních a základních modelů umělého neuronu vytvořili američtí vědci Warren McCulloch a Walter Pitts v roce 1943. Tento model zjednodušoval činnost tím, že neuron obsahoval excitační (synaptická váha $w = 1$) a inhibiční (synaptická váha $w = 0$) vazby. Neuron dále udržoval svůj vnitřní práh citlivosti, který, aby došlo k jeho excitaci, musel být překonán vnitřním potenciálem. Výstup toho neuronu tedy mohl být buď 0, nebo 1.

Další a jeden z nejpoužívanějších modelů je označován *perceptron*. Ten byl stvořen v roce 1957 Frankem Rosenblattem, který zobecnil model pánů McCullocha a Pittse. V tomto případě je potenciál určen váženým součtem vstupních signálů, a pokud překoná svůj práh citlivosti, nastane excitace neuronu, v opačném případě dojde k jeho inhibici. Potenciál je možné vyjádřit vztahem

$$\xi = \sum_{i=1}^n w_i x_i - \theta$$

kde w_i je synaptická váha, x_i vstupní signál a θ práh citlivosti. Výstupní odezva neuronu je definována pomocí aktivační funkce signum (funkce signum vrátí 1, je-li její argument větší než 0; v opačném případě vrátí 0):

$$y = \text{Sgn}(\xi)$$

Takto vytvořený perceptron je schopen vstupní data rozdělit do dvou skupin podle toho, je-li excitován ($y = 1$) nebo inhibován ($y = 0$) [5].

Dalším krokem je určení synaptických vah neuronu. K tomu poslouží dříve zmiňované trénovací množiny a učící algoritmus. Jeden takový algoritmus představil kanadský psycholog Donald Hebb a skládá se z několika kroků.

1. Nejprve je potřeba všechny váhy inicializovat náhodnými malými čísly, stejně se nastaví i práh citlivosti.
2. Síti je předložen vstupní vektor x z trénovací množiny a pomocí funkce $y = \text{Sgn}(\xi)$ se stanoví výstupní hodnota.
3. Výstupní hodnota je porovnávána s očekávanou hodnotou z trénovací množiny a nastává fáze adaptace vah, která může mít 3 stavy:
 - výstup sítě je shodný s očekávanou hodnotou $\rightarrow w_i(t+1) = w_i(t)$
 - výstup sítě je 0, očekáván byl 1 $\rightarrow w_i(t+1) = w_i(t) + x_i(t)$
 - výstup sítě je 1, očekáván byl 0 $\rightarrow w_i(t+1) = w_i(t) - x_i(t)$

Poslední část algoritmu lze ještě vylepšit *koefficientem učení* ($\eta, \eta \in \langle 0; 1 \rangle$), který slouží k ovlivnění razance změn synaptických vah během adaptace:

- výstup sítě je shodný s očekávanou hodnotou $\rightarrow w_i(t+1) = w_i(t)$
- výstup sítě je 0, očekáván byl 1 $\rightarrow w_i(t+1) = w_i(t) + \eta x_i(t)$
- výstup sítě je 1, očekáván byl 0 $\rightarrow w_i(t+1) = w_i(t) - \eta x_i(t)$

Tuto část můžeme zefektivnit přidáním *chyby odezvy* δ , která představuje rozdíl mezi očekávanou hodnotou ($d(t)$) a získanou hodnotou ($y(t)$):

$$\Delta = d(t) - y(t)$$

$$w_i(t+1) = w_i(t) + \eta \Delta x_i(t)$$

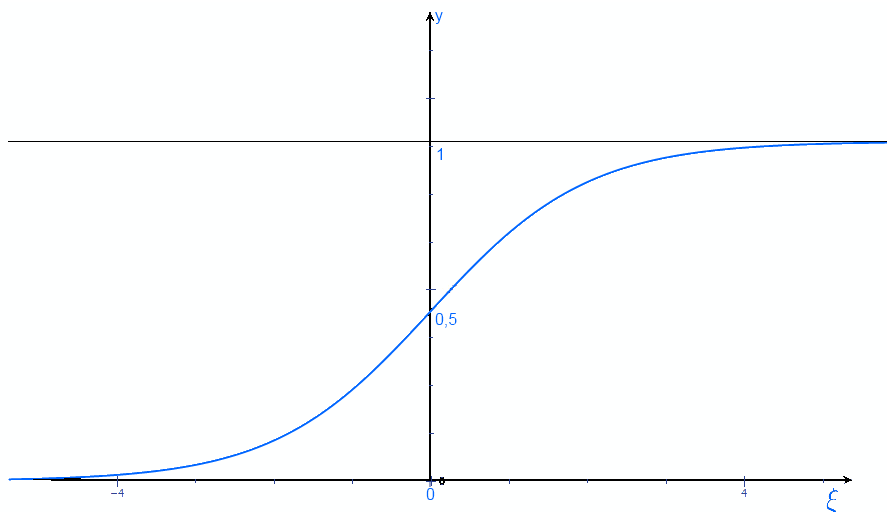
Další úpravu algoritmu učení řeší Widrow-Hoffovo pravidlo, kde se odezva a její chyba (delta pravidlo) nově vypočítává podle vztahu [5]:

$$y(t) = \sum_{i=0}^n w_i(t) x_i(t)$$

$$\Delta = d(t) - y(t) = d(t) - \sum_{i=0}^n w_i(t) x_i(t)$$

Nevýhodou takto definovaného perceptronu je to, že aktivační funkce neuronu je skoková a na výstupu tak umožňuje pouze dvě hodnoty. V případě, že aktivační funkci skokovou vyměníme za jinou, můžeme získat na výstupu neuronu značně více než dvě hodnoty. Mezi takové aktivační funkce můžou patřit [8]:

- sigmoidální funkce
- lineární funkce
- logaritmická funkce
- sinusová funkce
- funkce hyperbolické tangenty



Obrázek 9: Sigmoidální funkce

V tomto textu se budeme věnovat pouze sigmoidální funkci (sigmoidě, obrázek 9), která vypadá následovně:

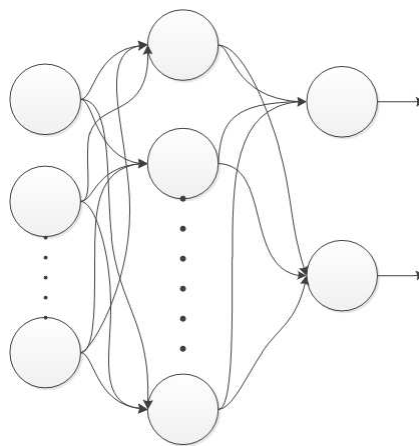
$$P(t) = \frac{1}{1 + e^{-t}}$$

a změni již definovanou rovnici výstupní odezvy neuronu na:

$$y = \frac{1}{1 + e^{-\lambda\xi}}$$

$$\xi = \sum_{i=0}^n w_i x_i$$

kde λ vyjadřuje strmost sigmoidu a ξ vnitřní potenciál neuronu. Sigmoidální aktivační funkce nám umožňuje vytvářet výstup z intervalu $\langle 0; 1 \rangle$. Nicméně ani takovýto model ještě není schopen řešit složité funkce, i když se vědělo, že je problém možné vyřešit dvouvrstvou neuronovou sítí, nebyl znám učící algoritmus pracující s více vrstvami. To bylo vyřešeno až v 80. letech po znovuobjevení algoritmu *backpropagation*, který je využíván dodnes.



Obrázek 10: Vícevrstvá neuronová síť

Vícevrstvá neuronová síť (obrázek 10) je složena vždy z alespoň tří vrstev neuronů – vstupní, vnitřní (skryté) a výstupní, přičemž vnitřních vrstev může být více (je ovšem dokázáno, že více vnitřních vrstev lze převést na jednu vnitřní vrstvu) a platí, že každý neuron z nižší vrstvy je spojen se všemi neurony vyšší vrstvy.

Nyní si popíšeme samotný adaptační algoritmus *backpropagation* (zpětné šíření), který má několik částí [5, 6, 7]:

1. Nejprve se vyřeší dopřední šíření signálu (ve směru od vstupní vrstvy do výstupní vrstvy, *feedforward*) – dojde k excitaci neuronů vstupní vrstvy a signál se rozešle neuronům následující vrstvy, pomocí synaptických vah se upraví. Každý neuron provede sumaci těchto signálů a pomocí své aktivační funkce je excitován. To se opakuje pro všechny vrstvy. Z výstupní vrstvy získáme odezvu sítě na konkrétní vstup.
2. Odezva sítě se porovná s očekávanou hodnotou z trénovací množiny. Rozdíl mezi těmito hodnotami je chyba sítě (Δ). Tato chyba se v určitém poměru (koeficient

učení η) zpětně propaguje (od výstupní vrstvy k vrstvě vstupní) do neuronové sítě přes úpravy synaptických vah. Cílem je snížit chybu Δ v další odezvě sítě.

3. Jakmile bude projita celá trénovací množina, vyhodnotí se chyba odezvy napříč touto množinou a pokud je vyšší, než naše očekávaná, celý algoritmus se opakuje od začátku.

Matematicky můžeme algoritmus backpropagation a jeho metody vyjádřit následujícími vztahy, kde hledáme minimum chybové funkce E pomocí úprav synaptických vah [5, 6]:

$$E = \frac{1}{2} \sum_{k=1}^p \sum_{j=1}^m (y_j - o_j)_k^2$$

$$\Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}} + \mu \Delta w'_{ij},$$

kde y_j je skutečná odezva j -tého neuronu na výstupu, o_j je požadovaná odezva ze vzoru trénovací množiny, p je celkový počet vzorů trénovací množiny, m je počet neuronů výstupní vrstvy. μ respektive Δw_{ij} značí koeficient vlivu změny vah respektive změnu váhy z předchozího kroku. Pro výpočet derivace z druhého vzorce, vyjdeme z následujících rovnic:

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial y} \cdot \frac{dy}{d\xi} \cdot \frac{\partial \xi}{\partial w_{ij}}$$

$$\frac{\partial \xi}{\partial w_{ij}} = x_i$$

$$\frac{dy}{d\xi} = y(1 - y)\lambda$$

Rovnice pro $\frac{\partial E}{\partial y}$ se bude lišit v tom, jestli je neuron součástí výstupní vrstvy nebo ne. V prvním případě bude pro vzor trénovací množiny k platit:

$$\frac{\partial E}{\partial y} = (y_j - o_j)_k$$

a v druhém:

$$\frac{\partial E}{\partial y} = \sum_{i=1}^m \delta^i \lambda y^i (1 - y^i) w^i$$

Algoritmus backpropagation je možné i dále vylepšovat (např. na parametrický backpropagation), stejně jako rozšiřovat možnosti perceptronu, nicméně už je to problematika mimo rámec této práce.

Pro předpovídání by bylo možné využít i jiné typy neuronových sítí a jejich varianty, mezi které lze zařadit:

- rekurentní vícevrstvá neuronová síť
- samoorganizující (Kohonenovy) mapy
- Hopfieldova síť

5 Predikce rychlostí vozidel

V minulé kapitole jsme popsali jednotlivé způsoby predikce, které nyní aplikujeme na dostupná dopravní data z projektu Floreon+ a stejně tak je i implementujeme. Výsledkem by měly být metody, které jsou schopny na základě reálných historických dat z provozu odhadnout vývoj rychlosti v budoucnu.

5.1 Aritmetický průměr dat

První metodou je vytvoření aritmetického průměru z dlouhodobých historických dat, které předcházejí hledanému dni a času. Tzn., chceme-li predikovat rychlost v 15.00 20. února 2011 (neděle), provedeme aritmetický průměr na naměřené rychlosti ze všech předcházejících nedělí v 15 hodin (13. února 2011, 6. února 2011, 30. ledna 2011, 23. ledna 2011,...). Vstupní data tedy mohou vypadat jako v tabulce 5 a výsledkem bude 62,83 km/h, což je předpovězená rychlost pro 20. 2. 2011 v 15.00. K této metodě bude

Datum	13. 2. 11	6. 2. 11	30. 1. 11	23. 1. 11	16. 1. 11	9. 1. 11
Rychlost	67	45	59	78	60	26
Datum	2. 1. 11	26. 12. 10	19. 12. 10	12. 12. 10	5. 12. 10	28. 11. 10
Rychlost	120	64	51	69	42	73

Tabulka 5: Vstupní data pro aritmetický průměr

náležet i její „vylepšená“ varianta, která bude aritmetický průměr počítat až z dat, ze kterých budou odstraněna odlehlá pozorování.

5.2 Lineární regrese

Lineární regrese bude navržena ve dvou variantách, které se liší pouze v tom, s jakými zdrojovými daty budou pracovat. První, označená jako *historická lineární regrese*, má stejný vstup, jako výše popsaná metoda s aritmetických průměrem, ale namísto průměru bude provádět aproximaci.

Druhá varianta, *aktuální lineární regrese*, pracuje pouze s rychlostmi, které přímo předcházejí času, ve kterém chceme predikovat rychlost vozidel. Chceme-li například předpovědět rychlost vozidel dnes v poledne, využije metoda údaje o rychlosti od dnešních jedenácti hodin do půlnoci předchozího dne (viz. tabulka 6). Výsledkem je hodnota 61,02 km/h, která udává předpokládanou rychlost vozidel v poledne.

Čas	11.00	10.00	9.00	8.00	7.00	6.00	5.00	4.00	3.00	2.00	1.00	0.00
Rychlost	67	45	59	78	60	26	120	64	51	69	42	73

Tabulka 6: Vstupní data pro aktuální lineární regresi

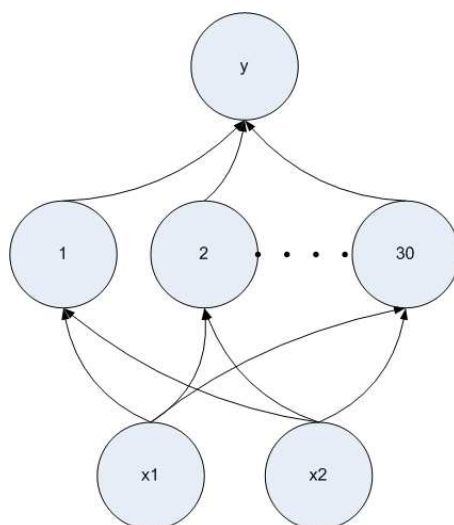
5.3 Neuronové sítě

Poslední metodou pro predikci rychlostí vozidel jsou neuronové sítě. Pro účely této práce budou navrženy tři modely ANN (lišící se v počtu vstupů, který bude od dvou do čtyř vstupů), které dostanou odpovídající počet (podle typu, dvě až čtyři hodiny) po sobě jdoucích hodinových rychlostí a jako výstup vrací rychlost vozidel v následující hodině (např. tabulka 7 obsahuje vstupy a výstupy pro neuronovou síť se dvěma vstupy). Neuronová síť bude naučena předpovídat rychlosti na základě dvou veličin – dne v týdnu (pondělí - neděle) a hodiny, pro kterou chceme znát předpověď.

vstup 1 11.00	vstup 2 12.00	výstup predikce pro 13.00
77 km/h	89 km/h	87 km/h
88 km/h	87 km/h	87 km/h
87 km/h	84 km/h	87 km/h

Tabulka 7: Příklad vstupů a výstupu neuronové sítě na úseku cesty 444720

5.3.1 Architektura a trénování neuronové sítě



Obrázek 11: Navržená neuronová síť

Tři typy navržených neuronových sítí budou složeny ze tří vrstev (vstupní, skrytá, výstupní), využijí adaptační algoritmu backpropagation a jako aktivační funkce poslouží sigmoidální funkce. Jak bylo stručně zmíněno výše, tyto tři typy ANN od sebe odlišuje počet vstupních neuronů, který začíná na dvou a končí na čtyřech neuronech. Počet neuronů vnitřní (skryté) vrstvy byl stanoven na třicet a výstupní vrstva obsahuje pouze

jediný neuron. Obrázek 11 znázorňuje návrh těchto neuronových sítí a používá variantu se dvěma vstupy (x_1, x_2). Dále zobrazuje skrytou a výstupní (y) vrstvu.

Aby bylo možné využívat neuronovou síť k předpovídání rychlostí vozidel, je nutné tuto síť naučit pomocí trénovací množiny. Ta se bude vytvářet z historických dopravních dat, jeden vzor trénovací množiny odpovídá jednomu řádku z tabulky 7 (v případě ANN se dvěma vstupy) a data se získávají zpětně vždy za určitý den týdne (stejně jako u metody s aritmetickým průměrem, akorát s několika hodinovými záznamy během dne).

Jelikož je proces učení neuronové sítě časově i výpočetně náročnější, budou se vytrénované sítě ukládat do databáze a v případě požadavku na predikci pomocí již dříve naučené sítě se místo nového trénování přímo stáhne z databáze a spustí s předanými hodnotami. Schéma tabulky je poměrně jednoduché s následujícími atributy:

Název atributu	Typ atributu	Příklad	Popis
NeuronNetId	Integer	1	ID záznamu
LinkId	Long	67898754	ID úseku cesty
NeuronNet	Binary		Serializovaný objekt ANN
HourOfPrediction	Integer	13	Hodina predikce (0-23)
DayOfWeek	Integer	5	Den predikce (1-7)
TimeAdded	DateTime	2. 3. 2010 4:45	Datum přidání záznamu

Tabulka 8: Schéma databáze pro ukládání ANN

6 Experimenty s prediktivními metodami

V této kapitole se zaměříme na otestování schopností predikce rychlostí s využitím metod představených v části *Predikce rychlostí vozidel* a popíšeme, jakých výsledků metody dosáhly.

K těmto experimentům bylo vybráno osmnáct úseků cest², které měly v době provádění více jak jedenáct tisíc záznamů (jedná se o úseky s největším počtem záznamů v databázi). Dále po provedení experimentů byly z výsledků odstraněny záznamy o rychlostech vozidel menších než 15 km/h, protože negativně zkreslovaly výstupní statistiku. Jakmile byla data připravena, spočítal se absolutní rozdíl mezi skutečnou zaznamenanou rychlostí a rychlostí předpovězenou. Tento rozdíl se poté vyjádřil v procentech vzhledem k reálné rychlosti a statisticky se porovnával tento procentuální údaj (např. predikovaná rychlost se od skutečné lišila o 12 %).

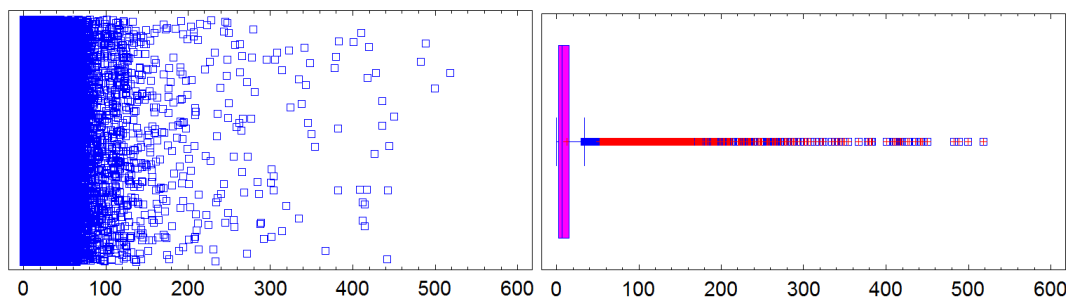
6.1 Aritmetický průměr dat

Pro vyzkoušení funkčnosti obou metod operujících s aritmetickým průměrem bylo vypočítáno 113 164 průměrů (pro každou metodu zvlášť) v období od 24. července 2008 do 29. ledna 2011. Tyto data byla ještě očištěna a ve výsledku se jednalo o 96 447 záznamů. Tabulka 9 poskytuje ve druhém sloupci úplný přehled o metodě aritmetického průměru a na obrázku 12 je zobrazen její bodový a krabicový graf. Z nich vyplývá, že se predikované rychlosti liší od reality v průměru o 12,6 % a $\frac{3}{4}$ záznamů se vejdu do rozdílu 15,2 %.

Charakteristika	Hodnoty	
	včetně odlehlých poz.	bez odlehlých poz.
Počet hodnot	96 447	96 447
Průměr	12,6368	12,2069
Medián	6,62454	5,82473
Rozptyl	385,219	382,839
Směrodatná odchylka	19,627	19,5663
Standardní chyba	0,0631989	0,0630034
Minimum	0,0	0,0
Maximum	518,652	518,652
Rozpětí	518,652	518,652
Dolní kvartil	2,76989	2,39232
Horní kvartil	15,2635	14,7335
Standardizovaná šikmost	11 606,4	11 338,0
Standardizovaná špičatost	6 158,12	5 922,57

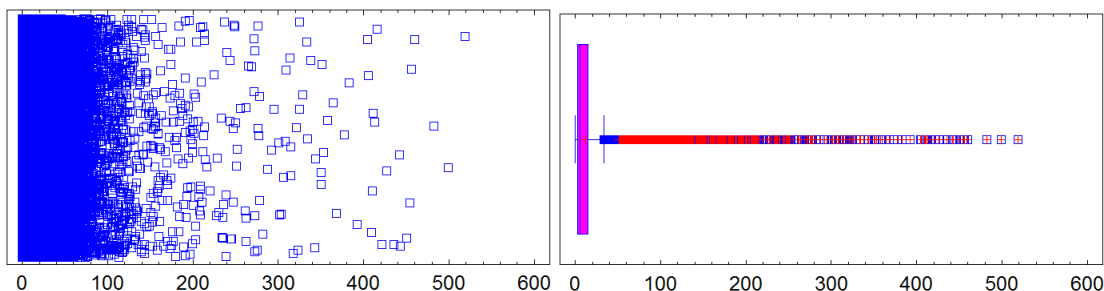
Tabulka 9: Základní statistika pro aritmetický průměr dat

²úseky 444720, 718596, 1092833, 394492, 989674, 661146, 806690, 1088899, 999118, 1030619, 85918, 412015, 1004168, 1005507, 730987, 772641, 1244309, 938633



Obrázek 12: Bodový a krabicový graf pro aritmetický průměr

Třetí sloupec tabulky 9 uvádí statistiku pro aritmetický průměr dat, která jsou očištěná o odlehlá (extrémní) pozorování. Ve srovnání s původní metodou poskytuje mírně lepší výsledky, průměrné hodnoty jsou menší o 0,4 procentního bodu. K dispozici je i bodový a krabicový graf (obrázek 13).



Obrázek 13: Bodový a krabicový graf pro aritmetický průměr dat bez odlehlých pozorování

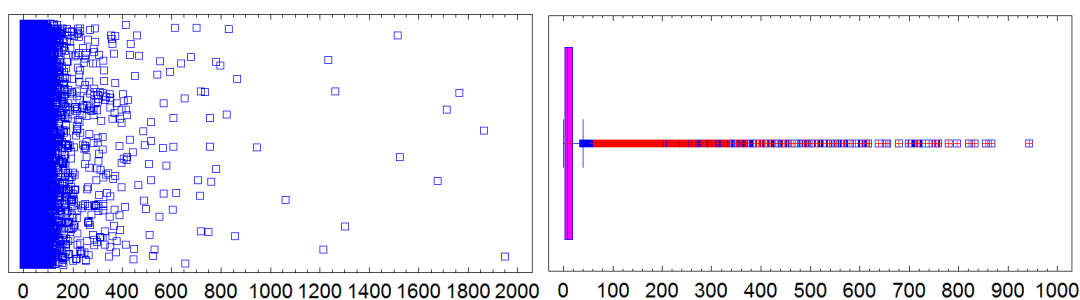
6.2 Lineární regrese

Stejně jako u aritmetického průměru dat jsou u lineární regrese dvě varianty. První varianta označená jako *historická lineární regrese*, má opět data z období od 24. července 2008 do 29. ledna 2011 a aproximovala 113 164 záznamů, které se po očištění zmenšily na 96 447 záznamů. Ze statistiky v tabulce 10 a bodového a krabicového grafu na obrázku 14 vyplynulo, že průměrný rozdíl od skutečné rychlosti činil 15,6 procent a $\frac{3}{4}$ záznamů měli tento rozdíl pod 17,9 %. Dále je z těchto grafů vidět, že metoda produkuje dost záznamů se značně velkým rozdílem od skutečné rychlosti.

Tento negativní jev je způsoben jednak malým množstvím dat, které jsou k dispozici a také tím, jak funguje samotná lineární regrese. Pokud má na vstupu k dispozici minimum dat (třeba 3 záznamy), která jsou poměrně daleko od predikovaného času a jsou ještě výrazně stoupající (nebo klesající), předpovídá rychlosti v řádu stovek kilometrů v hodině a to i v záporné variantě.

Charakteristika	Hodnoty	
	historická regrese	aktuální regrese
Počet hodnot	96 447	48 297
Průměr	15,6086	18,197
Medián	8,14655	10,092
Rozptyl	1 302,8	724,567
Směrodatná odchylka	36,0944	26,9178
Standardní chyba	0,116224	0,122484
Minimum	0,0	0,0
Maximum	3383,12	934,35
Rozpětí	3383,12	934,35
Dolní kvartil	3,43882	4,18017
Horní kvartil	17,9186	22,4442
Standardizovaná šikmost	51 795,8	-746,731
Standardizovaná špičatost	119 556,0	4 893,35

Tabulka 10: Základní statistika pro lineární regresi



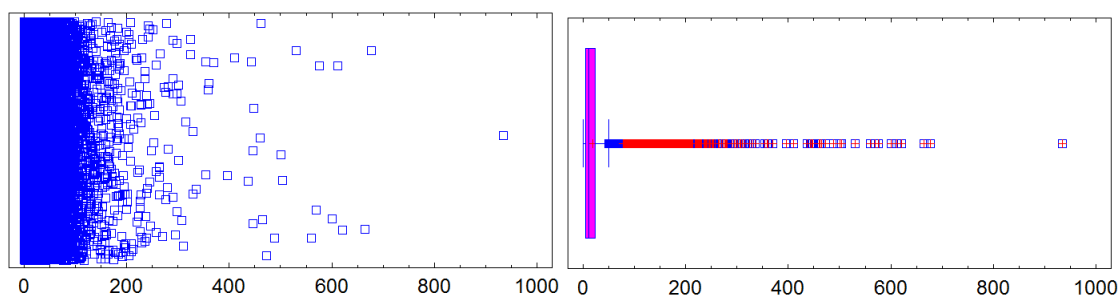
Obrázek 14: Bodový a krabicový graf pro lineární regresi z historických dat

Tabulka 10 také obsahuje hodnoty pro *aktuální lineární regresi*, která je na obrázku 15 doplněna o patřičné grafy. Tato varianta je postavena na záznamech od 2. ledna 2009 do 24. srpna 2010 v celkovém počtu 59 272 záznamů. Po očištění zbylo k dalšímu zkoumání 48 297 záznamů. Z těch je poznat, že metoda produkuje oproti *historické lineární regresi* horší výsledky. Průměrný rozdíl se zvýšil o 2,5 procentního bodu, u horního kvartilu se rozdíl mezi metodami zvednul už o 4,5 procentního bodu. Na druhou stranu nevytváří tolik záznamů s velmi výrazným rozdílem od skutečné rychlosti.

Bohužel i tato varianta je velmi závislá na dostupnosti dat, ze kterých se lineární regrese provádí a na spoustě úseků tak není schopná vracet žádné výsledky.

6.3 Neuronové sítě

Poslední metoda se věnuje neuronovým sítím. Pro všechny tři varianty byly použity data z rozmezí od července 2008 do března 2010. Z nich se vytvářely jednotlivé vzory trénovacích množin, které byly použity na učení neuronové sítě. Pro jednu konkrétní



Obrázek 15: Bodový a krabicový graf pro lineární regresi z aktuálních dat

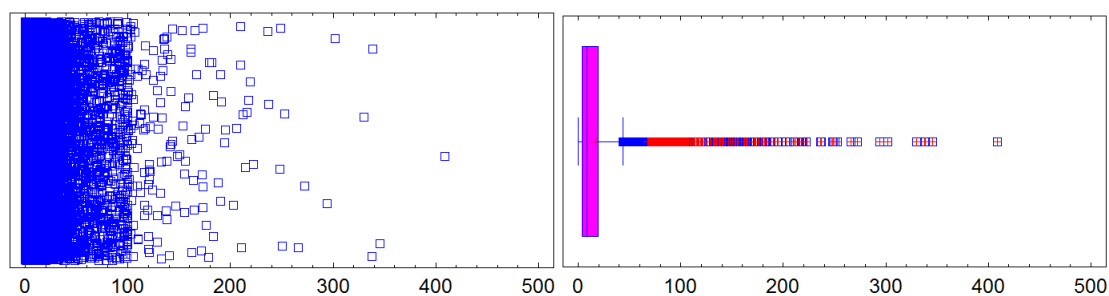
trénovací množinu jsou získány všechny historické rychlosti za jeden časový úsek v jeden den týdně během uvedeného období. Až byla neuronová síť naučena, provedlo se její testování a zhodnocení na datech od dubna 2010 do ledna 2011.

Charakteristika	Hodnoty		
	2 vstupy	3 vstupy	4 vstupy
Počet hodnot	14 559	11 278	8 974
Průměr	16,6429	17,6809	18,585
Medián	8,38164	8,49322	8,73655
Rozptyl	585,264	680,872	721,219
Směrodatná odchylka	24,1922	26,0935	26,8555
Standardní chyba	0,200498	0,245707	0,283492
Minimum	0,0000212211	0,000813898	0,000688196
Maximum	409,159	423,843	405,958
Rozpětí	409,159	423,842	405,957
Dolní kvartil	3,44953	3,49735	3,42089
Horní kvartil	19,5573	19,7833	21,1311
Standardizovaná šikmost	204,581	171,94	133,624
Standardizovaná špičatost	746,17	607,365	399,815

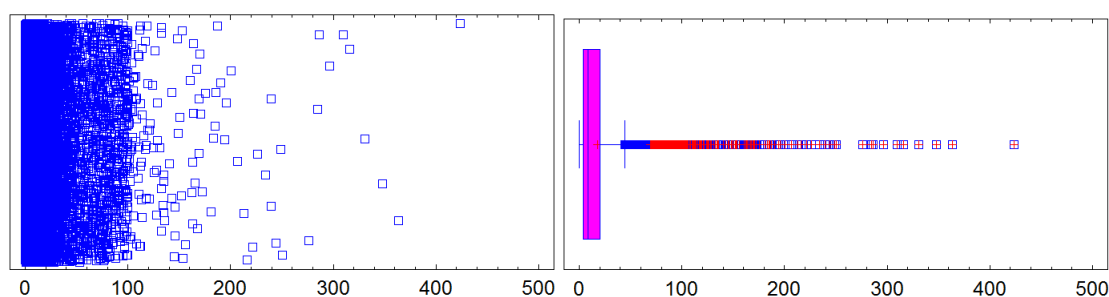
Tabulka 11: Základní statistika pro neuronovou síť se dvěma až čtyřmi vstupy

V tabulce 11 a obrázcích 16, 17 a 18 jsou získané údaje o výsledcích předpovídání pomocí ANN. Pro zhodnocení správnosti predikce byl použit celkem 48 501 záznam, z toho 19 502 (po očištění 14 559) záznamy pro neuronovou síť se dvěma vstupy, 15 823 (po očištění 11 278) pro síť se třemi vstupy a 13 176 (po očištění 8 974) záznamů pro síť se čtyřmi vstupy. Neuronová síť se čtyřmi vstupními neurony má nejmenší počet záznamů, protože je obtížné zajistit úseky, které by měly během jednoho dne dostatek po sobě jdoucích hodinových záznamů.

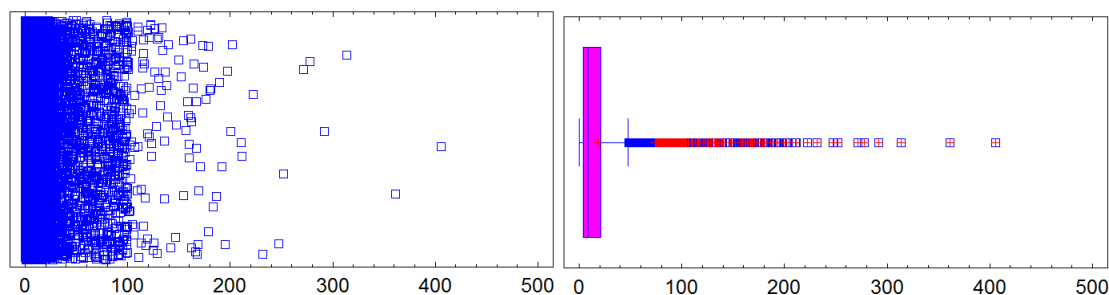
Ze srovnání vychází nejlépe síť se dvěma vstupy, která má průměrný rozdíl od reálných rychlostí 16,6 procent a $\frac{3}{4}$ záznamů mají rozdíl pod 19,6 procent. Zbylé dva typy sítí produkují vždy horší výsledky a ještě jsou zatíženy větší časovou a výkonovou náročností.



Obrázek 16: Bodový a krabicový graf pro neuronovou síť se dvěma vstupy



Obrázek 17: Bodový a krabicový graf pro neuronovou síť se třemi vstupy



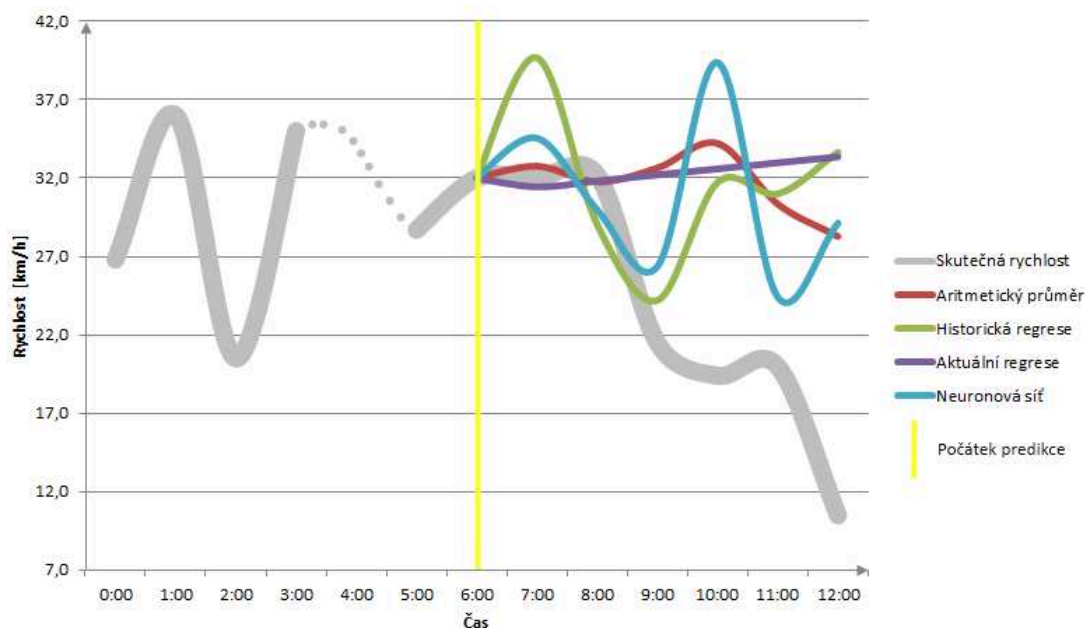
Obrázek 18: Bodový a krabicový graf pro neuronovou síť se čtyřmi vstupy

6.4 Reálné příklady predikce

Následující řádky budou na příkladech demonstrovat reálné výsledky předpovídání pomocí výše uvedených metod (aritmetický průměr bez extrémních pozorování, aktuální a historická lineární regrese, neuronová síť se dvěma vstupy). Grafy budou vždy obsahovat údaje za předpovídané období a pro porovnání bude přítomna i řada se skutečnou rychlostí včetně doby, která předchází prvnímu predikovanému záznamu.

6.4.1 Pondělí 14. března 2011, 7.00 – 12.00, úsek 989888

První oddíl se věnuje pondělnímu dopoledni na vybraném úseku cesty s ID 989888. Výsledek popisuje obrázek 19 a tabulka 12. Jak je z grafu vidět, skutečná rychlost za celé ráno a dopoledne hodně kolísala (dokonce i v noci) a žádná z předpovědí tento stav nepředpověděla. Pouze metoda založená na aritmetickém průměru historických dat dokázala vystihnout alespoň trend. To, že žádná z metod nedokázala predikovat poslední nízkou rychlost (10,5 km/h), způsobila absence podobných rychlostí v minulosti.



Obrázek 19: Graf s predikcí pro pondělí 14. března 2011

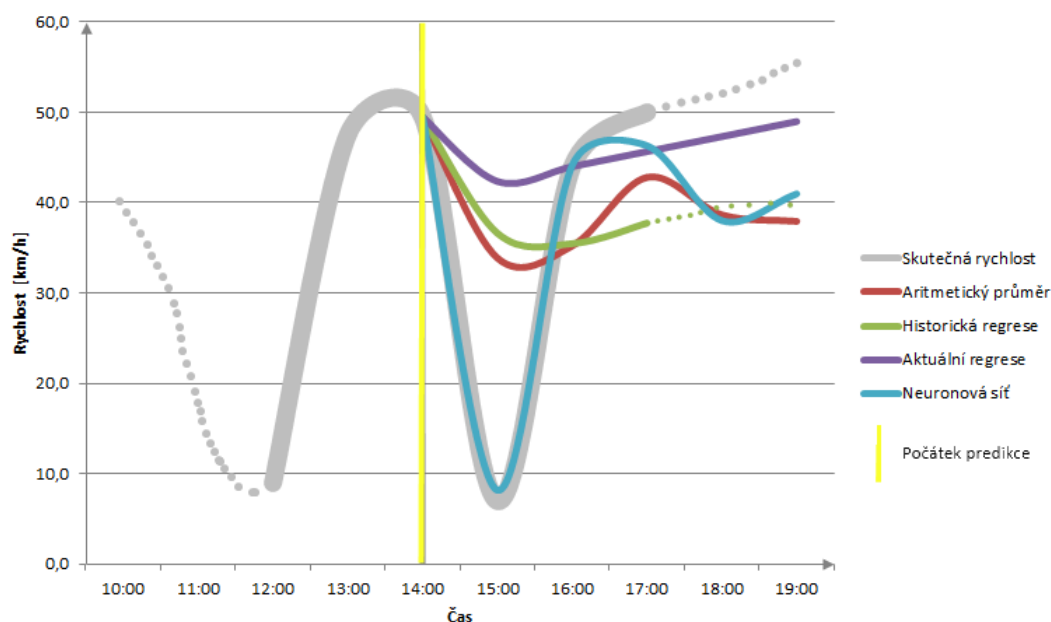
Čas	Skut. rychlost	Aritm. průměr	Hist. regrese	Akt. regrese	ANN
7.00	32	32,76	39,67	31,44	34,55
8.00	32,2	31,76	29,05	31,82	29,98
9.00	21,5	32,64	24,2	32,21	26,36
10.00	19,4	34,21	31,73	32,59	39,37
11.00	20,0	30,35	31	32,98	24,41
12.00	10,5	28,28	33,65	33,36	29,13

Tabulka 12: Tabulka předpovězených rychlostí pro pondělí 14. března 2011

6.4.2 Středa 23. března 2011, 15.00 – 19.00, úsek 230370

Středeční odpoledne na úseku cesty s ID 230370 poměrně dobře zachytila neuronová síť se dvěma vstupy, která z počátku v podstatě kopírovala skutečnou rychlost (obrázek

20). Z tabulky 13 je patrné, že ne vždy jsou k dispozici záznamy o skutečné rychlosti a z toho vyplývající následky v podobě metody bez výsledku (historická lineární regrese), protože ke svému výpočtu neměla dostatek vstupních dat. Ostatní metody predikce obstojně reprezentovaly alespoň trend rychlosti.



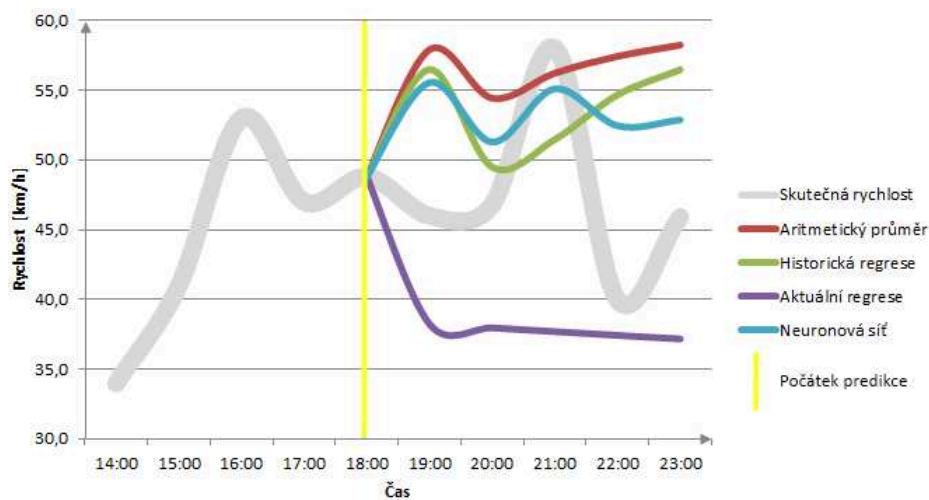
Obrázek 20: Graf s predikcí pro středu 23. března 2011

Čas	Skut. rychlost	Aritm. průměr	Hist. regrese	Akt. regrese	ANN
15.00	7,0	33,88	36,62	42,35	8,18
16.00	44,0	35,22	35,47	44	44,15
17.00	50,0	42,81	37,74	45,66	46,30
18.00		38,66		47,32	38,02
19.00	57,0	37,94	39,96	48,98	40,98

Tabulka 13: Tabulka předpovězených rychlostí pro středu 23. března 2011

6.4.3 Pátek, 11. března 2011, 19.00 – 23.00, úsek 989896

V páteční večer na úseku 989896 předpovídaly metody aritmetického průměru, historické lineární regrese a neuronové sítě téměř shodné výsledky, ale výkyv ve 22 hodin nezachytily (i když neuronová síť predikovala mírné snížení rychlosti), protože v dřívějších datech nebyl zaznamenán. Kompletní páteční výsledky popisuje obrázek 21 a tabulka 14.



Obrázek 21: Graf s predikcí pro pátek 11. března 2011

Čas	Skut. rychlost	Aritm. průměr	Hist. regrese	Akt. regrese	ANN
19.00	46,0	57,92	56,49	38,23	55,56
20.00	46,8	54,43	49,49	37,96	51,29
21.00	58,0	56,23	51,48	37,7	55,10
22.00	40,0	57,43	54,67	37,43	52,45
23.00	45,9	58,24	56,47	37,17	52,87

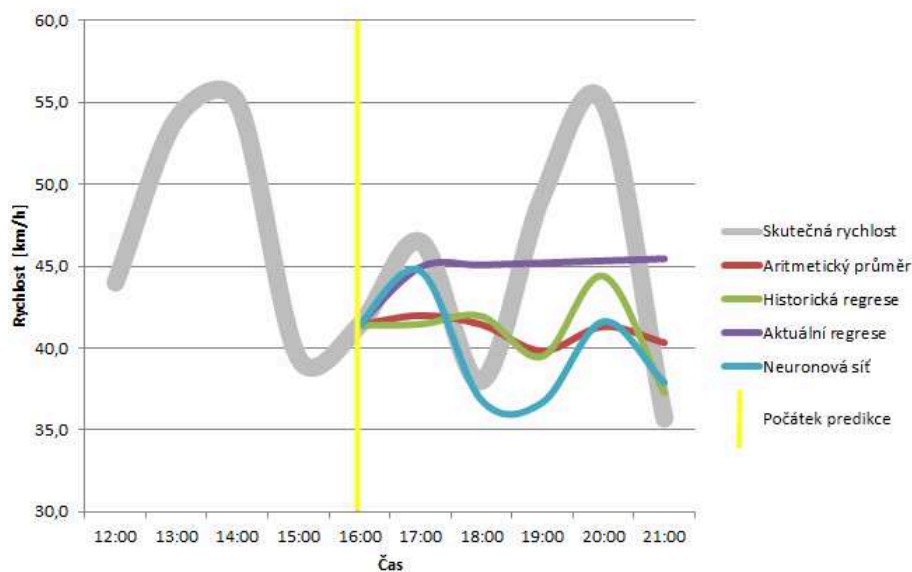
Tabulka 14: Tabulka předpovězených rychlostí pro pátek 11. března 2011

6.4.4 Neděle 13. března 2011, 17.00 – 21.00, úsek 546539

Nedělní graf (obrázek 22) pro úsek cesty s ID 546539 ukazuje, že vývoj asi nejlépe zachytila metoda neuronové sítě (ačkoliv nepředvídala ve dvacet hodin takový nárůst rychlosti), určitý trend kopírovala i historická lineární regrese. Přesná čísla nedělního večera má tabulka 15.

Čas	Skut. rychlost	Aritm. průměr	Hist. regrese	Akt. regrese	ANN
17.00	46,5	41,99	41,46	44,95	44,72
18.00	38,0	41,44	41,95	45,07	36,83
19.00	49,3	39,83	39,5	45,2	36,68
20.00	55,0	41,32	44,4	45,33	41,62
21.00	35,7	40,33	37,26	45,45	37,89

Tabulka 15: Tabulka předpovězených rychlostí pro neděli 13. března 2011



Obrázek 22: Graf s predikcí pro neděli 13. března 2011

6.5 Zhodnocení metod predikce

Experimenty prošlo dohromady skoro 400 tisíc záznamů a na jejich základě lze sestavit pořadí úspěšnosti jednotlivých metod (od nejlepší po nejhorší):

1. aritmetický průměr dat bez odlehlých pozorování
2. aritmetický průměr dat včetně odlehlých pozorování
3. historická lineární regrese
4. neuronová síť se dvěma vstupy
5. neuronová síť se třemi vstupy
6. neuronová síť se čtyřmi vstupy
7. aktuální lineární regrese

Jako nejvhodnější (nejpřesnější) pro predikci rychlostí vozidel se tedy jeví metoda založená na aritmetickém průměru historických dat, která jsou před zpracováním zbavena odlehlých pozorování. Tato metoda dokázala v průměru odhadnout rychlost s odchylkou 12,2 % od skutečné rychlosti. Výhodou metody je také fakt, že pro svou funkci nevyžaduje velké množství naměřených dat z minulosti.

Nedobře dopadly ve srovnání i neuronové sítě, přestože jejich výhodou je i to, že jsou imunní proti odlehlým pozorováním a jiným nestandardnostem a není tak potřeba je před trénováním sítě detekovat a odstraňovat. Neuronové sítě nejvíce doplatily na malé množství dostupných dopravních dat, ze kterých nebylo možné lépe trénovat síť. Dále se

ukázalo, že pro předpovídání dostačuje síť se dvěma vstupními neurony, která je nejméně závislá nejen na dostatku dat ale i čase. Jejich nevýhodou může být to, že proces učení neuronové sítě je výpočetně a časově náročnější a vzniká tak potřeba ukládat naučené sítě do databáze, aby nebylo nutné při každém požadavku znovu síť trénovat.

Lineární regrese se sice umístila mezi aritmetickým průměrem a ANN, ale obzvlášť varianta *aktuální lineární regrese* se jeví jako nejméně vhodná, protože velmi často není schopná určit výsledek kvůli nedostupnosti dat během sledovaného dne.

7 Implementace a propojení s webovou aplikací

Součástí této práce byla i implementace metod, které se věnují analýze dopravních dat a predikci rychlostí. Následující podkapitoly tak představí vývoj metod a jejich zapojení do většího celku.

7.1 Implementace

Nejprve byly vytvořeny obecné základní statisticko-analytické metody, které byly potřebné pro prediktivní metody:

- aritmetický průměr (i bez odlehlých (extrémních) pozorování)
- medián
- střední hodnota
- nejnižší a nejvyšší rychlosti
- směrodatná odchylka
- výběrový rozptyl
- zjišťování horních a dolních hradeb
- výběr a zpracování dnů v intervalu

Dalším krokem byla implementace metod pro predikci rychlostí vozidel a jejich otestování. Navrženy byly metody pro:

- lineární regresi
 - aktuální
 - historická
- aritmetický průměr dat (včetně varianty bez odlehlých pozorování)
- neuronovou síť
 - se dvěma vstupy
 - se třemi vstupy
 - se čtyřmi vstupy

Implementačně nejnáročnější prací bylo navrhnout predikování pomocí neuronových sítí [9]. Nejprve se musela vytvořit třída (*NewTrainingSet*), která zpracovala data získaná z databáze a přetvořila je do jednotlivých vzorů trénovací množiny. Protože neuronová síť pracuje se sigmoidální aktivační funkcí, jsou všechny rychlosti vozidel převáděny (výpis 1) z a do rozsahu $< 0; 1 >$.

```

static internal double ConvertToAnotherRange(double number, double minIn, double maxIn,
double minOut, double maxOut)
{
    return ((number - minIn) * (maxOut - minOut) / (maxIn - minIn) + minOut);
}

```

Výpis 1: Převod rychlostí z a do rozsahu < 0; 1 >

Takto vytvořená trénovací množina je předána neuronové síti (třída *NeuralNetwork*), kde poslouží k jejímu učení. Neuronová síť pro svoji architekturu, učení a získávání výsledků používá knihovnu *NeuronDotNet* [8]. Ve výpisu 2 je znázorněno vytvoření neuronové sítě se dvěma vstupními neurony, 30 neurony ve skryté vrstvě a jedním výstupním neuronem.

```

public BackpropagationNetwork Network { private get; set; }
NeuronCount = 30;
TrainingSpeed = 0.3d;
NumberOfInputs = 2;
NumberOfOutputs = 1;
public void CreateNetwork()
{
    LinearLayer inputLayer = new LinearLayer(NumberOfInputs);
    SigmoidLayer hiddenLayer = new SigmoidLayer(NeuronCount);
    SigmoidLayer outputLayer = new SigmoidLayer(NumberOfOutputs);

    new BackpropagationConnector(inputLayer, hiddenLayer).Initializer = new RandomFunction(0d,
        0.3d);
    new BackpropagationConnector(hiddenLayer, outputLayer).Initializer = new RandomFunction(0d,
        0.3d);
    Network = new BackpropagationNetwork(inputLayer, outputLayer);
    Network.SetLearningRate(TrainingSpeed);
}

```

Výpis 2: Vytvoření neuronové sítě

Po vytvoření sítě se předá vytvořená trénovací množina a spustí se proces učení. Až bude neuronová síť naučena, zavoláním metody *RunNetwork* s patřičnými argumenty dojde ke spuštění sítě, která k předaným hodnotám spočítá odpovídající výstup.

```

EpochCount = 30000;
public void NetworkTrain(NewTrainingSet ts)
{
    TrainingSet trainingSet = ts.CreateTrainingSet(out minSpeedRange, out maxSpeedRange);
    Network.Learn(trainingSet, EpochCount);
}

public double[] RunNetwork(double[] input)
{
    double[] output = Network.Run(input);
    double[] editedOutput = new double[output.Length];
    for (int i = 0; i < output.Length; i++)
    {
        editedOutput[i] = ConvertToAnotherRange(output[i], 0, 1, minSpeedRange, maxSpeedRange);
    }
}

```

```

return editedOutput;
}

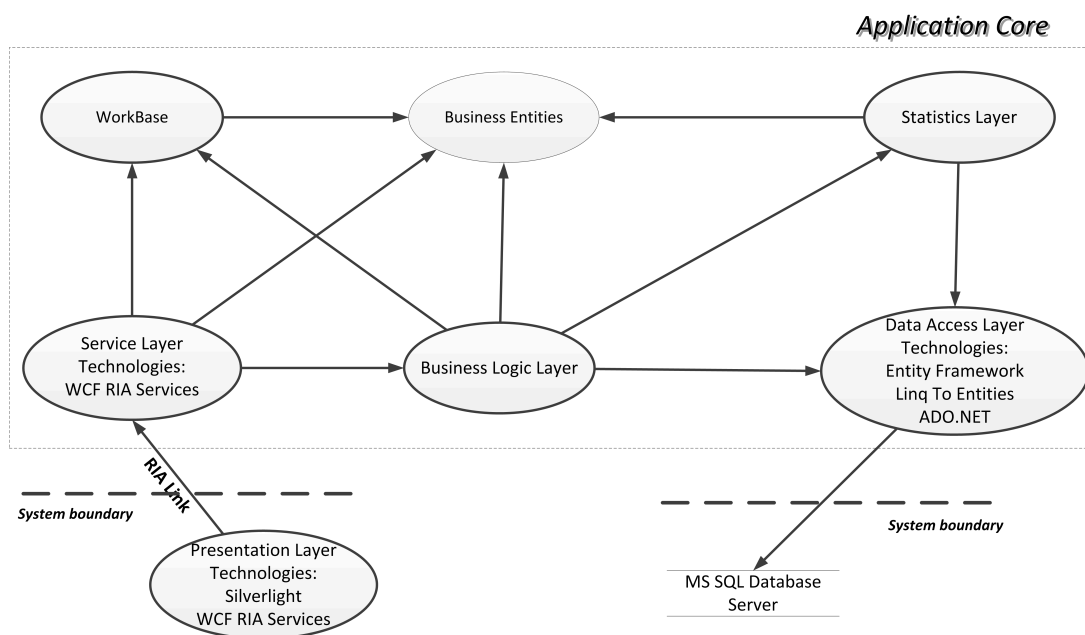
```

Výpis 3: Učení a spuštění neuronové sítě

Metody byly v rámci experimentů otestovány pomocí vytvořeného programu, který stahoval naměřená data z databáze a po zpracování je předával k výpočtům jednotlivým technikám předpovídání a jejich výsledky porovnával se skutečně naměřenými rychlostmi. Jakmile byl dokončen vývoj a zhodnocení jednotlivých prostředků predikce, začala integrace vytvořených metod do souhrnného projektu.

7.2 Propojení

Jelikož je tato práce součástí rozsáhlejšího systému, na kterém se podílí i další kolegové, bylo by na místě zde zveřejnit její zapojení do tohoto celku. Primární spolupráce směřovala k práci, které se zabývala především architekturou a propojením celého systému a měla tak dopad i na implementační část této práce. Ta reprezentuje vrstvu *Statistic Layer* na obrázku 23 s návrhem systému.



Obrázek 23: Architektura celého systému

Statistická vrstva poskytuje zbytku systému několik analytických metod:

- základní statistika (průměry, minima, maxima, rozptyly,...)
- historická data (zaznamenané rychlosti vozidel v minulosti)
- predikce rychlostí

- pomocí neuronových sítí
- pomocí aritmetického průměru dat
- pomocí aritmetického průměru dat bez odlehlých (extrémních) pozorování
- pomocí aktuální lineární regrese
- pomocí historické lineární regrese

Ty byly navrženy s ohledem na maximální jednoduchost použití, jak je vidět z následujícího výpisu 4:

```

public static BasicStatistics GetBasicStatistics(long routeld, DateTime dateFrom, DateTime
    dateTo);
public static Dictionary<DateTime, double> GetHistoricalData(long routeld, DateTime dateFrom,
    DateTime dateTo);
public enum PredictionType { NeuralNetwork, HistoricalData, EnhancedHistoricalData,
    ActualLinearRegression, HistoricalLinearRegression };
public static Dictionary<DateTime, double> GetPrediction(PredictionType predictionType, long
    routeld, DateTime dateFrom, DateTime dateTo);

BasicStatistics bs = Statistics .GetBasicStatistics(706780134, new DateTime(2011, 03, 01), new
    DateTime(2011, 03, 03));
Dictionary<DateTime, double> history = Statistics.GetHistoricalData(706780134, new DateTime
    (2011, 03, 01), new DateTime(2011, 03, 31));
Dictionary<DateTime, double> prediction = Statistics.GetPrediction(PredictionType.NeuralNetwork
    , 592662078, new DateTime(2010, 12, 10, 8, 0, 0), new DateTime(2010, 12, 10, 12, 0, 0));

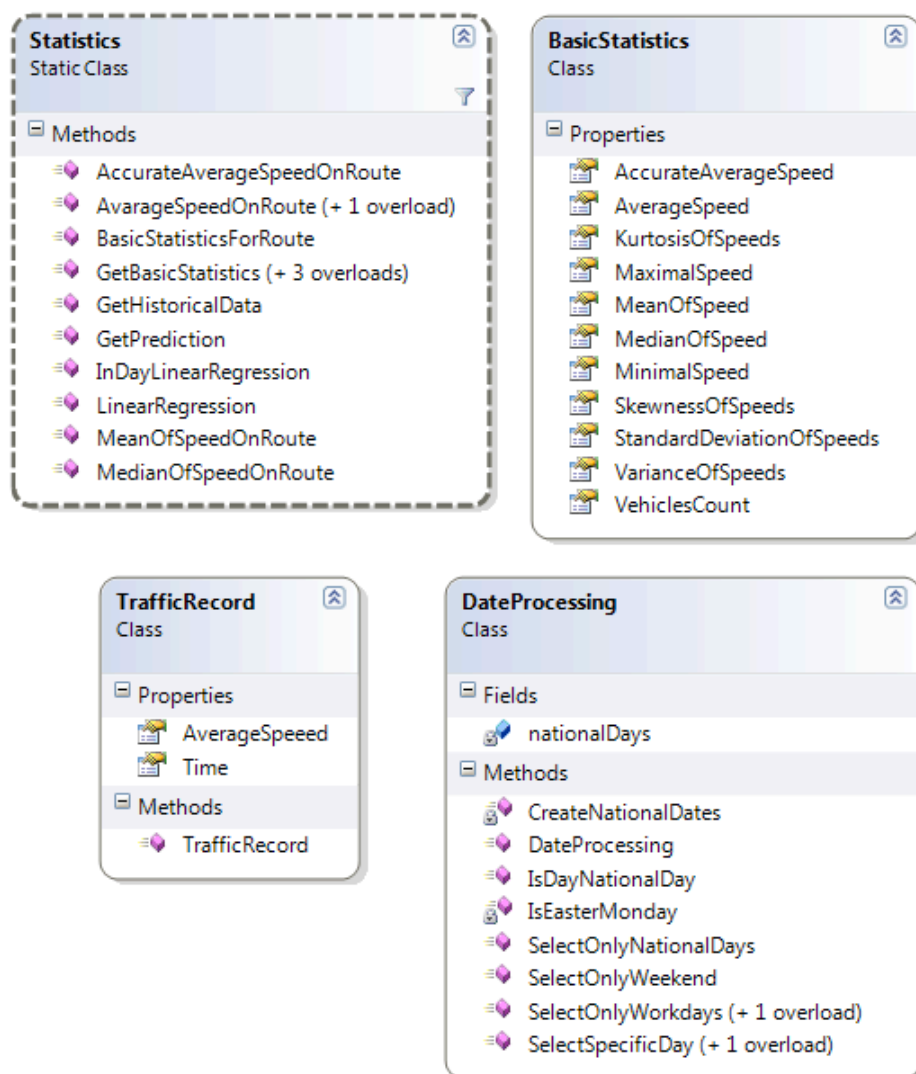
```

Výpis 4: Definice a použití metod třídy Statistics

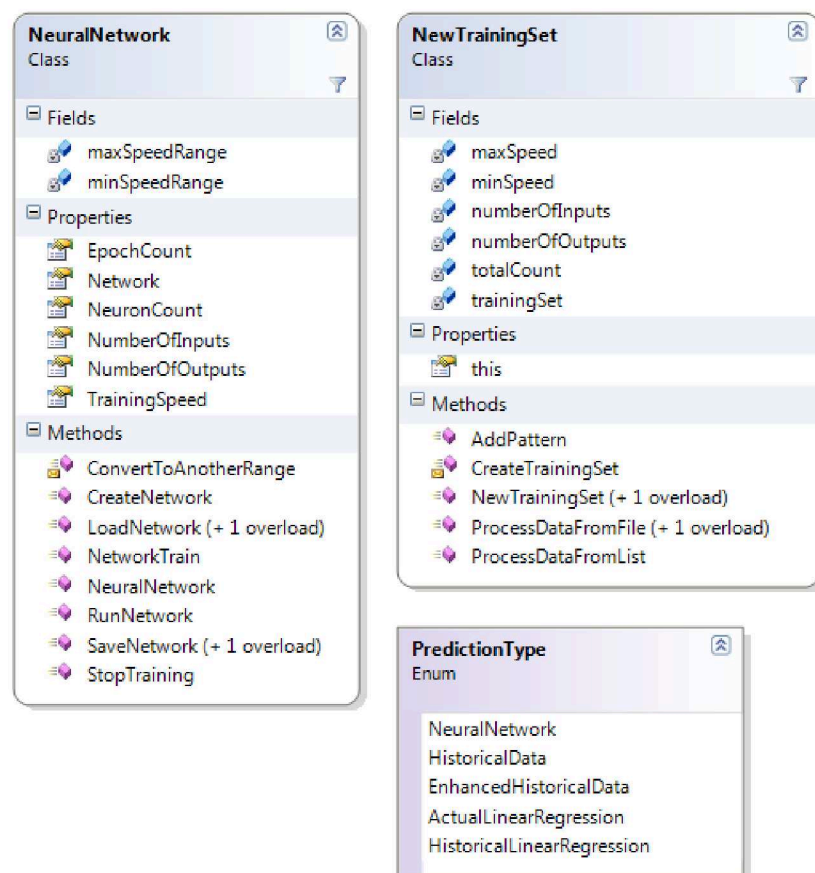
Obrázky 24 a 25 představují navržené třídy části StatisticLayer využívané pro práci s daty a pro statistické dotazování dalších částí systému.

7.3 Webová aplikace

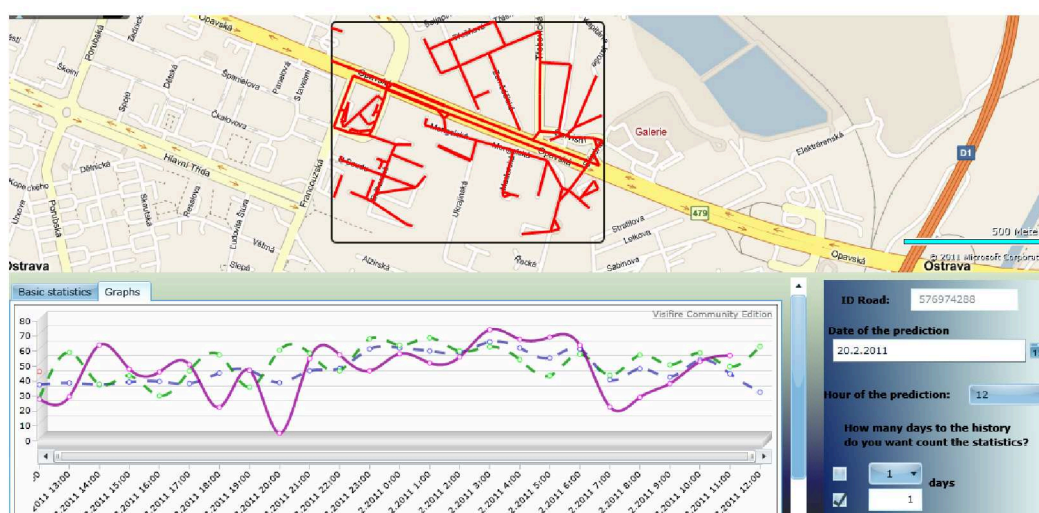
Kromě samotného návrhu architektury se práce kolegy věnovala i webové aplikaci, která umožňuje z mapy vybírat úseky cest a k nim zobrazovat mimo jiné údaje získané právě z analytických metod. Uživatel, který s aplikací pracuje, si může vybrat, pro jaké časové období chce provádět analýzu či predikci a výsledky z metod se vykreslí do grafu, jak je patrné z obrázku 26.



Obrázek 24: Navržené třídy v rámci projektu StatisticLayer



Obrázek 25: Navržené třídy v rámci projektu StatisticLayer



Obrázek 26: Webová aplikace používající předpovídání rychlosti

8 Vize do budoucna

Tato práce představila některé analytické a prediktivní techniky, které lze v budoucnu zcela jistě vylepšovat. Základem je dostupnost kvalitních dopravních dat v co největším množství, protože to je největší slabina současného systému. V kapitole popisující sběr dopravních dat, byly popsány některé způsoby, o které by bylo vhodné systém rozšířit. Jelikož smyčkové detektory a mýtné brány zatím nejsou příliš rozšířené, mohly by se do sběru dopravních dat zapojit dopravní kamery, které jsou dnes dostupné na frekventovaných křižovatkách a pomocí analýzy zachyceného obrazu počítat množství aut a v omezené míře i jejich rychlost.

Mohl by se také rozšířit počet spolupracujících aut (plovoucích vozidel), které posílají svá data o rychlosti a poloze, ideálně se zapojením nefiremních vozidel, případně tato data získat od třetích stran (např. společnost Telefónica O_2 disponuje více jak 30 000 sledovanými auty [16]). Ostatní techniky sběru nejsou do vyřešení legislativních problémů aktuální.

Kvalitnější dopravní data zajisté přinesou lepší výsledky předpovídání rychlosti vozidel, neboť dnes existuje velký počet úseků, kde neexistence nebo malý počet dat znemožňuje nasazení prediktivních metod. Tyto metody by si také zasloužily vylepšování především výkonu. Nejvíce je to patrné u neuronových sítí, kde v podstatě není možné jejich nasazení v reálném čase, protože pro každý úsek cesty, pro každou hodinu a pro každý den v týdnu je potřeba získat dopravní data, z nich vytvořit trénovací množinu a pomocí ní naučit neuronovou síť. Tento proces je časově a výkonnostně velmi náročný a na běžných počítačích může trvat trénování neuronové sítě pomocí jedné trénovací množiny až 30 sekund s plným využitím procesoru.

Jelikož se dá učení neuronových sítí paralelizovat, stojí za zvážení využití *High-performance computing*, což znamená použití superpočítačů a počítačových klastrů, na kterých by měl být proces učení podstatně kratší a ANN by se daly používat k predikci i v reálném čase. Po úspěšném zprovoznění superpočítače na VŠB-TU v rámci projektu *IT4Innovations* by bylo vhodné vyzkoušet na něm jeho skutečné možnosti při trénování neuronových sítí.

Výsledky predikce rychlosti by se daly vylepšit i filtrací vstupních dat, která by mohla obsahovat informace o dění v oblasti umístění silnice (úseku). Tyto informace mohou být následující:

- počasí
 - mlha
 - námraza
 - sníh
 - déšť
 - bez problému
- dopravní omezení
 - dopravní nehoda

- uzavírka silnice
 - oprava cesty
 - omezení rychlosti
- akce (s rozlišením velikosti)
 - společenské
 - kulturní
 - sportovní

Na jejich základě by prediktivní metody používaly tato filtrovaná data a mohly by přesněji určit rychlost vozidel v případě, že nastane některá z vyjmenovaných situací. Alternativně by metody pracovaly s běžnými daty, ale po předpovězení rychlosti by tento výsledek mohl být upraven nějakým koeficientem vytvořeným na základě výše uvedených situací, např. při námraze sníží výslednou rychlost o 20 procent.

9 Závěr

Cílem této práce bylo popsat získávání a zpracování reálných dopravních dat. Nejprve tedy bylo nutné se seznámit s existujícími dopravními systémy a zjistit, jak lze dopravní data získávat. Ze systému Floreon+ Traffic byla získávána dopravní data poskytovaná smluvními vozidly (plovoucí vozidla), která se zpracovala pro další využití – analýzu dat a predikci rychlostí. Po hlubším zkoumání dat bylo také zjištěno, že ve spoustě případů neposkytují dostatečné informace, které bychom potřebovali. Zejména jde o úplnou absenci hodnot ve sledovaném období případně o nedostatečný počet vozidel vytvářející hodinové průměrné rychlosti.

Pro provádění predikce rychlostí vozidel bylo nutné se seznámit s vybranými technikami předpovídání, jednalo se o aritmetický průměr, lineární regresi a neuronové sítě. Získané poznatky byly realizovány do několika metod, které se využily při experimentech. Během nich se testovaly schopnosti jednotlivých technik předpovídat rychlosti vozidel.

Na závěr byly analytické a prediktivní metody zpracovány tak, aby byly jednoduše použitelné v rámci celého vyvíjeného systému. Metody se použily v práci kolegy, který vytvářel další část systému a to konkrétně ve webové aplikaci umožňující uživatelům vybírat z mapy jednotlivé úseky cest a na nich si zobrazovat odhadnuté rychlosti vozidel ve zvoleném období, případně ještě základní statistické údaje o vybraném úseku.

Poslední kapitola práce se věnovala možnému vylepšování předpovídání rychlostí v budoucnu. Jelikož je největší problém v dostupnosti či spíše v nedostupnosti vhodných dopravních dat, měl by se systém doplnit o jejich další zdroje. První možností je rozšířit interní flotilu spolupracujících vozidel případně získat údaje od společností, které je komerčně nabízejí. Druhou možností je analýza obrazů získaných z dopravních kamer. Jiné možnosti nejsou v brzké době realizovatelné. Predikci rychlostí by bylo možné zpřesnit i rozšířením spektra informací vstupních dat o údaje o počasí, o dopravní omezení nebo o různé pořádané akce.

Přínosy této práce by se dali shrnout do několika následujících bodů:

- získání statistických informací o konkrétním úseku cesty
- srovnání vybraných prediktivních metod s ohledem na existující dopravní data
- možnost předpovídání rychlostí vozidel na jednotlivých úsecích cesty v určitou dobu

Mým osobním přínosem při řešení této diplomové práce byla možnost vyzkoušet si práci v týmu se všemi pozitivními i negativními vlastnostmi na rozsáhlejší projekt. Získal jsem spoustu nových informací o tom, jakým způsobem se získávají dopravní data a jak je lze využít k různým účelům v dopravních systémech.

10 Reference

- [1] BRIŠ, Radim; LITSCHMANNOVÁ, Martina. *Statistika I. : pro kombinované studium* [online]. Ostrava : VŠB - Technická univerzita Ostrava, 2004 [cit. 2011-03-15]. Dostupné z WWW: <<http://homel.vsb.cz/zot005/lit/statistika-pr-mat.html>>.
- [2] OTIPKA, Petr; ŠMAJSTRLA, Vladislav. *Pravděpodobnost a statistika* [online]. Ostrava : VŠB - Technická univerzita Ostrava, [2010] [cit. 2011-03-15]. Dostupné z WWW: <<http://homen.vsb.cz/oti73/cdpast1>>. ISBN 80-248-1194-4.
- [3] ZVÁRA, Karel. *Regrese*. Praha : Matfyzpress, 2008. 253 s. ISBN 978-80-7378-041-8.
- [4] HALTUFOVÁ, Zuzana. *Zpracování a analýza dopravních dat z reálného provozu*. [s.l.], 2010. 44 s. Diplomová práce. VŠB - Technická univerzita Ostrava, Fakulta elektrotechniky a informatiky.
- [5] VONDRÁK, Ivo. *Neuronové sítě* [online]. Ostrava : VŠB - Technická univerzita Ostrava, 2009 [cit. 2011-03-28]. Dostupné z WWW: <http://vondrak.cs.vsb.cz/download/Neuronove_site.pdf>.
- [6] VOLNÁ, Eva. *Vícevrstvá neuronová síť jako univerzální aproximátor*. Automatizace. Listopad 2009, 52, 11, s. 658-660. Dostupný také z WWW: <<http://www.automatizace.cz/article.php?a=2606>>.
- [7] TUČKOVÁ, Jana. *Vybrané aplikace umělých neuronových sítí při zpracování signálů*. Praha : Česká technika - nakladatelství ČVUT, 2009. 224 s. ISBN 978-80-01-04229-8.
- [8] *NeuronDotNet - Artificial Neural Networks in C#* [online]. c2011 [cit. 2011-02-19]. Dostupný z WWW: <<http://neurondotnet.freehostia.com>>.
- [9] HEATON, Jeff. *Introduction to Neural Networks for C#*. St. Louis : Heaton Research, Inc, 2008. 428 s. ISBN 1-60439-009-3.
- [10] Český rozhlas [online]. 2010-09-06 [cit. 2011-04-02]. *Naše vysílání i pro vaše navigace... a víte kudy kam!*. Dostupné z WWW: <http://www.rozhlas.cz/zelenavlna/rds-tmc/_zprava/780204>.
- [11] Dopravní info [online]. c2009 [cit. 2011-04-02]. *Jednotný systém dopravních informací pro ČR (JSDI)*. Dostupné z WWW: <<http://portal.dopravniinfo.cz/jsdi>>.
- [12] Floreon [online]. c2006 [cit. 2011-04-02]. *Informace o projektu*. Dostupné z WWW: <<http://floreon.vsb.cz/Web/About/cz/Baseinfo.aspx>>.
- [13] MARTINOVIČ, Jan, et al. *Floreon : modulární systém pro podporu rozhodování v krizovém řízení*. In Sborník sympozia GIS Ostrava 2009. Ostrava : VŠB - Technická univerzita Ostrava, 2009. s. 1-6. Dostupné z WWW: <http://gis.vsb.cz/GIS_Ostrava/GIS_Ova_2009/sbornik/Lists/Papers/109.pdf>. ISBN 978-80-87294-00-0.

-
- [14] RADECKÝ, Michal. *Floreon+ Traffic : Systém pro modelování, simulace a monitorování dopravy*. IT Systems. 2009, 9, s. 42-45. Dostupný také z WWW: <http://www.it4i.cz/files/media.itsystems_0909.pdf>. ISSN 1802-615X.
- [15] SDT [online]. Praha : 2011-03-01 [cit. 2011-04-09]. *Dopravní data: nutná podmínka efektivního dopravního systému v ČR*. Dostupné z WWW: <http://www.sdt.cz/download/doc/Dopravni_data_final_040311.pdf>.
- [16] LINHART, Filip. *Dopravní informace v reálném čase založené na datech z plovoucích vozidel*. In How communication technology can support road traffic management and increase road safety in cities [online]. Praha : [CVIS], 2010-03-10 [cit. 2011-04-10]. Dostupné z WWW: <http://www.cvisproject.org/download/Presentations/Prague/2_cvis_training_traffic_information.pdf>.